GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 17:34:32 ; Search time 3055.3 Seconds
(without alignments)
3801.338 Million cell updates/sec

Title: US-09-727-769A-5
Perfect score: 555
Sequence: 1 ttggcgagtgtaattcctga.....atgtatccagctgtggattt 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	555	100.0	555	6	AX113621	AX113621 Sequence
	2	555	100.0	1080	6	AX113623	AX113623 Sequence
	3	555	100.0	1380	1	AB046594	AB046594 Chryseoba
	4	344.2	62.0	555	6	AR159957	AR159957 Sequence
	5	344.2	62.0	555	6	E59337	E59337 Novel prote
	6	344.2	62.0	1080	6	AR159962	AR159962 Sequence
	7	344.2	62.0	1080	6	E59342	E59342 Novel prote
	8	43	7.7	158254	2	AC098659	AC098659 Rattus no
c	9	41	7.4	125020	9	AF429315	AF429315 Homo sapi
c	10	39.8	7.2	148000	2	AC094081	AC094081 Homo sapi
	11	38.4	6.9	196932	2	AC100728	AC100728 Mus muscu
c	12	38	6.8	27068	3	CEC15A7	Z67735 Caenorhabdi
c	13	38	6.8	185947	9	AC092630	AC092630 Homo sapi
	14	37.8	6.8	161729	5	AL590155	AL590155 Zebrafish
	15	37.6	6.8	161221	9	AC022390	AC022390 Homo sapi
c	16	37.4	6.7	58282	2	AC106326	AC106326 Rattus no
c	17	37.2	6.7	30836	3	U67948	U67948 Caenorhabdi
c	18	37.2	6.7	132749	2	AC006778	AC006778 Caenorhab
	19	37.2	6.7	165373	2	AC093916	AC093916 Homo sapi
	20	37	6.7	155344	9	AC026407	AC026407 Homo sapi
c	21	37	6.7	181946	2	AC097330	AC097330 Pan trogl
	22	37	6.7	182974	2	AC105148	AC105148 Homo sapi
	23	36.8	6.6	162371	9	AC016687	AC016687 Homo sapi
	24	36.8	6.6	173201	9	AC008169	AC008169 Homo sapi
c	25	36.6	6.6	95566	9	AL161714	AL161714 Human DNA
	26	36.6	6.6	154457	2	AC099239	AC099239 Rattus no
	27	36.6	6.6	155842	2	AC025650	AC025650 Homo sapi
c	28	36.6	6.6	161479	8	AC090482	AC090482 Genomic s
c	29	36.6	6.6	177829	2	AC096872	AC096872 Rattus no
	30	36.6	6.6	252360	2	AC091272	AC091272 Mus muscu
	31	36.4	6.6	132977	2	RN510D20	AL603723 Rattus no
c	32	36.4	6.6	204122	2	AL671867	AL671867 Mus muscu
c	33	36.2	6.5	1725	10	AB040609	AB040609 Rattus no
c	34	36.2	6.5	172792	9	AC062031	AC062031 Homo sapi
	35	36.2	6.5	187445	30	AC024639	AC024639 Homo sapi
c	36	36	6.5	19576	6	AX251958	AX251958 Sequence
c	37	36	6.5	19576	6	AX349007	AX349007 Sequence
c	38	36	6.5	37170	9	AL356054	AL356054 Human DNA
	39	36	6.5	68822	2	AC109141	AC109141 Mus muscu
c	40	36	6.5	68822	2	AC109141	AC109141 Mus muscu
	41	36	6.5	179598	9	CNS01DRO	AL117693 Human chr
	42	35.8	6.5	23342	3	CBRG26J16	AC084538 Caenorhab
c	43	35.6	6.4	2489	1	AF112369	AF112369 Ehrlichia
	44	35.6	6.4	102126	2	AC098049	AC098049 Rattus no
c	45	35.6	6.4	170948	2	AC016931	AC016931 Homo sapi

ALIGNMENTS

RESULT 1
AX113621
LOCUS AX113621 555 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 5 from Patent EP1106696.
ACCESSION AX113621
VERSION AX113621.1 GI:13939799
KEYWORDS .
SOURCE Chryseobacterium sp. No. 9670.
ORGANISM Chryseobacterium sp. No. 9670
Bacteria; CFB group; Flavobacteria; Flavobacteriaceae;
Chryseobacterium.
REFERENCE 1 (bases 1 to 555)
AUTHORS Yamaguchi,S.
TITLE Protein-deamidating enzyme, microorganism producing the same, gene encoding the same, production process therefor, and use thereof
JOURNAL Patent: EP 1106696-A 5 13-JUN-2001;
Amano Enzyme Inc. (JP)
FEATURES Location/Qualifiers
source 1. .555
/organism="Chryseobacterium sp. No. 9670"
/db_xref="taxon:161487"
BASE COUNT 166 a 119 c 107 g 163 t
ORIGIN

```
  Query Match           100.0%;  Score 555;  DB 6;  Length 555;
  Best Local Similarity  100.0%;  Pred. No. 6.7e-138;
  Matches  555;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTGGCGAGTGTAATTCCTGATGTAGCTACATTAAATTCTTTATTCAATCAAATAAAGAAT 60

Qy     61 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGTCTTGCGGTACCTCTACGGCGTCCTCACCATGCATCACATTCAGATATCCTGTAGAC 120

Qy    121 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGATGTTATGCAAGAGCCCATAAGATGAGACAAATCTTAATGAACAACGGCTATGACTGT 180

Qy    181 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAAAAACAATTTGTATACGGAAACCTAAAGGCATCAACAGGAACTTGCTGTGTGGCGTGG 240

Qy    241 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCTACCACGTTGCAATATTGGTAAGCTATAAAAATGCTTCCGGAGTAACGGAAAAAAGA 300

Qy    301 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATTATTGATCCTTCACTATTTTCAAGCGGTCCTGTAACAGATACAGCATGGAGAAACGCT 360

Qy    361 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCGTTAACACCTCTTGCGGATCTGCATCCGTTTCCTCTTATGCTAATACTGCAGGAAAT 420

Qy    421 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTTTATTACAGAAGTCCTAGTAATTCTTACCTGTATGACAACAATCTGATCAATACCAAC 480

Qy    481 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGTGTACTGACTAAATTTTCACTGCTTTCCGGATGTTCTCCTTCACCTGCACCGGATGTA 540

Qy    541 tccagctgtggattt 555
          |||||||||||||||
Db    541 TCCAGCTGTGGATTT 555


RESULT   2
AX113623
LOCUS       AX113623               1080 bp    DNA     linear   PAT 30-MAY-2001
DEFINITION  Sequence 7 from Patent EP1106696.
ACCESSION   AX113623
VERSION     AX113623.1  GI:13939800
KEYWORDS    .
SOURCE      Chryseobacterium sp. No. 9670.
  ORGANISM  Chryseobacterium sp. No. 9670
            Bacteria; CFB group; Flavobacteria; Flavobacteriaceae;
            Chryseobacterium.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Yamaguchi,S.
  TITLE     Protein-deamidating enzyme, microorganism producing the same, gene
            encoding the same, production process therefor, and use thereof
  JOURNAL   Patent: EP 1106696-A 7 13-JUN-2001;
            Amano Enzyme Inc. (JP)
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Chryseobacterium sp. No. 9670"
                     /db_xref="taxon:161487"
     CDS             61. .1023
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC37857.1"
                     /db_xref="GI:13939801"
                     /translation="MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKD
                     FGKTVPVGIDEENGMIKVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKP
                     NSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGT
                     STASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYH
                     VAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNV
                     YYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCGF"
     mat_peptide     466. .1020
                     /product="unnamed"
BASE COUNT      371 a    204 c    203 g    302 t
ORIGIN


  Query Match           100.0%;  Score 555;  DB 6;  Length 1080;
  Best Local Similarity  100.0%;  Pred. No. 6.6e-138;
  Matches  555;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    466 TTGGCGAGTGTAATTCCTGATGTAGCTACATTAAATTCTTTATTCAATCAAATAAAGAAT 525

Qy     61 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    526 CAGTCTTGCGGTACCTCTACGGCGTCCTCACCATGCATCACATTCAGATATCCTGTAGAC 585

Qy    121 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    586 GGATGTTATGCAAGAGCCCATAAGATGAGACAAATCTTAATGAACAACGGCTATGACTGT 645

Qy    181 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    646 GAAAAACAATTTGTATACGGAAACCTAAAGGCATCAACAGGAACTTGCTGTGTGGCGTGG 705

Qy    241 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    706 AGCTACCACGTTGCAATATTGGTAAGCTATAAAAATGCTTCCGGAGTAACGGAAAAAAGA 765

Qy    301 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    766 ATTATTGATCCTTCACTATTTTCAAGCGGTCCTGTAACAGATACAGCATGGAGAAACGCT 825

Qy    361 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    826 TGCGTTAACACCTCTTGCGGATCTGCATCCGTTTCCTCTTATGCTAATACTGCAGGAAAT 885

Qy    421 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    886 GTTTATTACAGAAGTCCTAGTAATTCTTACCTGTATGACAACAATCTGATCAATACCAAC 945

Qy    481 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    946 TGTGTACTGACTAAATTTTCACTGCTTTCCGGATGTTCTCCTTCACCTGCACCGGATGTA 1005

Qy    541 tccagctgtggattt 555
          |||||||||||||||
Db   1006 TCCAGCTGTGGATTT 1020


RESULT   3
AB046594
LOCUS       AB046594               1380 bp    DNA     linear   BCT 27-APR-2001
DEFINITION  Chryseobacterium proteolyticum prgA gene for protein-glutaminase,
            complete cds.
ACCESSION   AB046594
VERSION     AB046594.1  GI:12597204
KEYWORDS    .
SOURCE      Chryseobacterium proteolyticum (strain:9670) DNA.
  ORGANISM  Chryseobacterium proteolyticum
            Bacteria; CFB group; Flavobacteria; Flavobacteriaceae;
            Chryseobacterium.
REFERENCE   1  (sites)
  AUTHORS   Yamaguchi,S., Jeenes,D.J. and Archer,D.B.
  TITLE     Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme
```

```
           that deamidates glutaminyl residues in proteins Purification,
           characterization and gene cloning
  JOURNAL  Eur. J. Biochem. 268 (5), 1410-1421 (2001)
  MEDLINE  21153247
REFERENCE  2  (bases 1 to 1380)
  AUTHORS  Yamaguchi,S.
  TITLE    Direct Submission
  JOURNAL  Submitted (26-JUL-2000) Shotaro Yamaguchi, Amano Pharmaceutical Co.
           Ltd., Research and Development; Suei, Kagamigahara, Gifu 509-0108,
           Japan (E-mail:LDV01447@nifty.ne.jp, Tel:81-583-79-1220,
           Fax:81-583-79-1232)
FEATURES             Location/Qualifiers
     source          1. .1380
                     /organism="Chryseobacterium proteolyticum"
                     /strain="9670"
                     /db_xref="taxon:118127"
     sig_peptide     181. .243
                     /gene="prgA"
     CDS             181. .1143
                     /gene="prgA"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-glutaminase"
                     /protein_id="BAB21508.1"
                     /db_xref="GI:12597205"
                     /translation="MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKD
                     FGKTVPVGIDEENGMIKVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKP
                     NSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGT
                     STASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYH
                     VAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNV
                     YYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCGF"
     gene            181. .1143
                     /gene="prgA"
     mat_peptide     586. .1140
                     /gene="prgA"
                     /evidence=experimental
                     /product="unnamed"
BASE COUNT      472 a    247 c    247 g    414 t
ORIGIN
```

Query Match 100.0%; Score 555; DB 1; Length 1380;
Best Local Similarity 100.0%; Pred. No. 6.5e-138;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 60
Db     586 TTGGCGAGTGTAATTCCTGATGTAGCTACATTAAATTCTTTATTCAATCAAATAAAGAAT 645

Qy      61 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 120
Db     646 CAGTCTTGCGGTACCTCTACGGCGTCCTCACCATGCATCACATTCAGATATCCTGTAGAC 705

Qy     121 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 180
Db     706 GGATGTTATGCAAGAGCCCATAAGATGAGACAAATCTTAATGAACAACGGCTATGACTGT 765

Qy     181 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 240
Db     766 GAAAAACAATTTGTATACGGAAACCTAAAGGCATCAACAGGAACTTGCTGTGTGGCGTGG 825

Qy     241 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 300
Db     826 AGCTACCACGTTGCAATATTGGTAAGCTATAAAAATGCTTCCGGAGTAACGGAAAAAAGA 885

Qy     301 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 360
Db     886 ATTATTGATCCTTCACTATTTTCAAGCGGTCCTGTAACAGATACAGCATGGAGAAACGCT 945

Qy     361 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 420
Db     946 TGCGTTAACACCTCTTGCGGATCTGCATCCGTTTCCTCTTATGCTAATACTGCAGGAAAT 1005

Qy     421 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 480
Db    1006 GTTTATTACAGAAGTCCTAGTAATTCTTACCTGTATGACAACAATCTGATCAATACCAAC 1065

Qy     481 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 540
Db    1066 TGTGTACTGACTAAATTTTCACTGCTTTCCGGATGTTCTCCTTCACCTGCACCGGATGTA 1125

Qy     541 tccagctgtggattt 555
Db    1126 TCCAGCTGTGGATTT 1140
```

```
RESULT   4
AR159957
LOCUS       AR159957                555 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 5 from patent US 6251651.
ACCESSION   AR159957
VERSION     AR159957.1  GI:16222831
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Yamaguchi,S. and Matsuura,A.
  TITLE     Protein-deamidating enzyme, gene encoding the same, production
            process therefor, and use thereof
  JOURNAL   Patent: US 6251651-A 5 26-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="unknown"
BASE COUNT      159 a    122 c    112 g    162 t
ORIGIN
```

Query Match 62.0%; Score 344.2; DB 6; Length 555;
Best Local Similarity 76.7%; Pred. No. 1.7e-81;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy       7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
Db       7 AGTGTTATTCCTGATCTGGCAACGCTGAACAGTTTATTTACCCAGATCAAAAACCAGGCT 66

Qy      67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
Db      67 TGCGGAACTTCTACAGCATCTTCTCCTTGTATCACCTTCAGATATCCGGTTGACGGATGT 126

Qy     127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
Db     127 TATGCAAGGGCTCACAAAATGAGACAAATCCTATTGAACGCCGGCTATGACTGTGAAAAG 186

Qy     187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
Db     187 CAGTTCGTATATGGTAATCTGAGAGCTTCTACAGGAACATGCTGTGTATCATGGGTATAT 246

Qy     247 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 306
Db     247 CACGTAGCAATTTTGGTAAGCTTCAAAAATGCTTCAGGAATTGTTGAAAAAAGAATCATA 306

Qy     307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
Db     307 GATCCTTCATTATTCTCCAGCGGTCCTGTAACAGATTCTGCATGGAGAGCTGCATGTACC 366

Qy     367 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
Db     367 AACACAAGCTGCGGATCTGCGTCTGTATCTTCCTACGCCAATACAGCAGGAAATGTTTAC 426

Qy     427 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 486
Db     427 TACAGAAGTCCGTCAGGTTCATTACTGTATGATAACAACTATGTGAATACCAATTGTGTA 486

Qy     487 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 546
```

```
Db    487 TTAAACATATTCTCATCCCTTTCAGGATGTTCTCCTTCCCCAGCACCAAGTGTAGCAAGC 546

Qy    547 tgtggattt 555
          |||||||||
Db    547 TGTGGATTT 555


RESULT   5
E59337
LOCUS       E59337                    555 bp    DNA     linear   PAT 07-FEB-2001
DEFINITION  Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof.
ACCESSION   E59337
VERSION     E59337.1  GI:13023304
KEYWORDS    JP 2000050887-A/3.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Shotaro,Y.A.M.M.
  TITLE     Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof
  JOURNAL   Patent: JP 2000050887-A 3 22-FEB-2000;
            AMANO PHARMACEUT CO LTD
COMMENT     OS   Chryseobacterium gleum
            PN   JP 2000050887-A/3
            PD   22-FEB-2000
            PF   04-JUN-1999 JP 1999158703
            PR
            PI   SHOTARO YAMAGUCHI,AKIRA MATSUURA
            PC   C12N15/09,C12N9/80//A21D2/36,A23C9/154,A23C11/06,A23J3/00, PC
             A23J3/10,
            PC   A23J3/16,A23J3/18,A23J3/34,A23L1/176,A23L1/23,A23L1/317, PC
            (C12N9/80,C12R1:01),(C12N9/80,C12R1:20),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   source          1. .555
            FT                   /organism='Chryseobacterium gleum'.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      159 a    122 c    112 g    162 t
ORIGIN


  Query Match             62.0%;  Score 344.2;  DB 6;  Length 555;
  Best Local Similarity   76.7%;  Pred. No. 1.7e-81;
  Matches  421;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy      7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
          ||||| |||||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db      7 AGTGTTATTCCTGATCTGGCAACGCTGAACAGTTTATTTACCCAGATCAAAAACCAGGCT 66

Qy     67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
          ||||| || ||||| || || || || || ||||| |||||||||| || |||||||||
Db     67 TGCGGAACTTCTACAGCATCTTCTCCTTGTATCACCTTCAGATATCCGGTTGACGGATGT 126

Qy    127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    127 TATGCAAGGGCTCACAAAATGAGACAAATCCTATTGAACGCCGGCTATGACTGTGAAAAG 186

Qy    187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    187 CAGTTCGTATATGGTAATCTGAGAGCTTCTACAGGAACATGCTGTGTATCATGGGTATAT 246

Qy    247 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 306
          ||||| ||||| |||||||||| |||||||||| ||| |    |||||||||| ||
Db    247 CACGTAGCAATTTTGGTAAGCTTCAAAAATGCTTCAGGAATTGTTGAAAAAAGAATCATA 306

Qy    307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
          ||||||||| |||| || |||||||||||||||||| | |||||||||   || ||
Db    307 GATCCTTCATTATTCTCCAGCGGTCCTGTAACAGATTCTGCATGGAGAGCTGCATGTACC 366

Qy    367 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
          |||||    |||||||||||| || || || || || || ||||| ||||||||||||||
Db    367 AACACAAGCTGCGGATCTGCGTCTGTATCTTCCTACGCCAATACAGCAGGAAATGTTTAC 426

Qy    427 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 486
          ||||||||||       ||| |  |||||||| |||||      | ||||||||| ||||||
Db    427 TACAGAAGTCCGTCAGGTTCATTACTGTATGATAACAACTATGTGAATACCAATTGTGTA 486

Qy    487 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 546
          | |  | ||| |||   ||||| |||||||||||||| || |||||   |||| | |||
Db    487 TTAAACATATTCTCATCCCTTTCAGGATGTTCTCCTTCCCCAGCACCAAGTGTAGCAAGC 546

Qy    547 tgtggattt 555
          |||||||||
Db    547 TGTGGATTT 555


RESULT   6
AR159962
LOCUS       AR159962                1080 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 11 from patent US 6251651.
ACCESSION   AR159962
VERSION     AR159962.1  GI:16222841
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Yamaguchi,S. and Matsuura,A.
  TITLE     Protein-deamidating enzyme, gene encoding the same, production
            process therefor, and use thereof
  JOURNAL   Patent: US 6251651-A 11 26-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="unknown"
BASE COUNT      354 a    210 c    205 g    311 t
ORIGIN


  Query Match             62.0%;  Score 344.2;  DB 6;  Length 1080;
  Best Local Similarity   76.7%;  Pred. No. 1.6e-81;
  Matches  421;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy      7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
          ||||| |||||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db    469 AGTGTTATTCCTGATCTGGCAACGCTGAACAGTTTATTTACCCAGATCAAAAACCAGGCT 528

Qy     67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
          ||||| || ||||| || || || || || ||||| |||||||||| || |||||||||
Db    529 TGCGGAACTTCTACAGCATCTTCTCCTTGTATCACCTTCAGATATCCGGTTGACGGATGT 588

Qy    127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    589 TATGCAAGGGCTCACAAAATGAGACAAATCCTATTGAACGCCGGCTATGACTGTGAAAAG 648

Qy    187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    649 CAGTTCGTATATGGTAATCTGAGAGCTTCTACAGGAACATGCTGTGTATCATGGGTATAT 708

Qy    247 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 306
          ||||| ||||| |||||||||| |||||||||| ||| |    |||||||||| ||
Db    709 CACGTAGCAATTTTGGTAAGCTTCAAAAATGCTTCAGGAATTGTTGAAAAAAGAATCATA 768

Qy    307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
          ||||||||| |||| || |||||||||||||||||| | |||||||||   || ||
Db    769 GATCCTTCATTATTCTCCAGCGGTCCTGTAACAGATTCTGCATGGAGAGCTGCATGTACC 828

Qy    367 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
          |||||    |||||||||||| || || || || || || ||||| ||||||||||||||
Db    829 AACACAAGCTGCGGATCTGCGTCTGTATCTTCCTACGCCAATACAGCAGGAAATGTTTAC 888

Qy    427 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 486
          ||||||||||       ||| |  |||||||| |||||      | ||||||||| ||||||
```

```
Db    889 TACAGAAGTCCGTCAGGTTCATTACTGTATGATAACAACTATGTGAATACCAATTGTGTA 948

Qy    487 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 546
          | |  | ||| |||   ||||| ||||||||||||||| || |||||   |||| | |||
Db    949 TTAAACATATTCTCATCCCTTTCAGGATGTTCTCCTTCCCCAGCACCAAGTGTAGCAAGC 1008

Qy    547 tgtggattt 555
          |||||||||
Db   1009 TGTGGATTT 1017


RESULT   7
E59342
LOCUS       E59342                  1080 bp    DNA     linear   PAT 07-FEB-2001
DEFINITION  Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof.
ACCESSION   E59342
VERSION     E59342.1  GI:13023309
KEYWORDS    JP 2000050887-A/8.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Shotaro,Y.A.M.M.
  TITLE     Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof
  JOURNAL   Patent: JP 2000050887-A 8 22-FEB-2000;
            AMANO PHARMACEUT CO LTD
COMMENT     OS   Chryseobacterium gleum
            PN   JP 2000050887-A/8
            PD   22-FEB-2000
            PF   04-JUN-1999 JP 1999158703
            PR
            PI   SHOTARO YAMAGUCHI,AKIRA MATSUURA
            PC   C12N15/09,C12N9/80//A21D2/36,A23C9/154,A23C11/06,A23J3/00, PC
             A23J3/10,
            PC   A23J3/16,A23J3/18,A23J3/34,A23L1/176,A23L1/23,A23L1/317, PC
            (C12N9/80,C12R1:01),(C12N9/80,C12R1:20),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (61). .(1020)
            FT   mat_peptide     (463). .(1017).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      354 a    210 c    205 g    311 t
ORIGIN


  Query Match           62.0%;  Score 344.2;  DB 6;  Length 1080;
  Best Local Similarity   76.7%;  Pred. No. 1.6e-81;
  Matches  421;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy      7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
          ||||| |||||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db    469 AGTGTTATTCCTGATCTGGCAACGCTGAACAGTTTATTTACCCAGATCAAAAACCAGGCT 528

Qy     67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
          ||||| || ||||| || || || || || || ||||| |||||||||| || |||||||||
Db    529 TGCGGAACTTCTACAGCATCTTCTCCTTGTATCACCTTCAGATATCCGGTTGACGGATGT 588

Qy    127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    589 TATGCAAGGGCTCACAAAATGAGACAAATCCTATTGAACGCCGGCTATGACTGTGAAAAG 648

Qy    187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    649 CAGTTCGTATATGGTAATCTGAGAGCTTCTACAGGAACATGCTGTGTATCATGGGTATAT 708

Qy    247 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 306
          ||||| ||||| |||||||||| |||||||||| ||| |    |||||||||| ||
Db    709 CACGTAGCAATTTTGGTAAGCTTCAAAAATGCTTCAGGAATTGTTGAAAAAAGAATCATA 768
```

```
Qy    307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
          ||||||||| |||| || |||||||||||||||||| | |||||||||   || ||
Db    769 GATCCTTCATTATTCTCCAGCGGTCCTGTAACAGATTCTGCATGGAGAGCTGCATGTACC 828

Qy    367 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
          |||||    |||||||||| || || || || || || ||||| ||||||||||||||
Db    829 AACACAAGCTGCGGATCTGCGTCTGTATCTTCCTACGCCAATACAGCAGGAAATGTTTAC 888

Qy    427 tacagaagtcctagtaatttcttacctgtatgacaacaatctgatcaataccaactgtgta 486
          ||||||||||      ||| |  |||||||| |||||     | ||||||| ||||||
Db    889 TACAGAAGTCCGTCAGGTTCATTACTGTATGATAACAACTATGTGAATACCAATTGTGTA 948

Qy    487 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 546
          | |  | ||| |||   ||||| ||||||||||||||| || |||||   |||| | |||
Db    949 TTAAACATATTCTCATCCCTTTCAGGATGTTCTCCTTCCCCAGCACCAAGTGTAGCAAGC 1008

Qy    547 tgtggattt 555
          |||||||||
Db   1009 TGTGGATTT 1017


RESULT   8
AC098659
LOCUS       AC098659               158254 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-100L14, *** SEQUENCING IN PROGRESS
            ***, 77 unordered pieces.
ACCESSION   AC098659
VERSION     AC098659.3  GI:17973414
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 158254)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
```

```
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158254)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:17064362.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHVY
                Center clone name: CH230-100L14
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 121530 bases at least Q40
                Consensus quality: 129199 bases at least Q30
                Consensus quality: 134968 bases at least Q20
                Estimated insert size: 109535; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 1.5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 77 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5298: contig of 5298 bp in length
            *     5299     5398: gap of unknown length
            *     5399     9928: contig of 4530 bp in length
            *     9929    10028: gap of unknown length
            *    10029    14124: contig of 4096 bp in length
            *    14125    14224: gap of unknown length
            *    14225    18693: contig of 4469 bp in length
            *    18694    18793: gap of unknown length
            *    18794    22503: contig of 3710 bp in length
            *    22504    22603: gap of unknown length
            *    22604    24527: contig of 1924 bp in length
            *    24528    24627: gap of unknown length
            *    24628    28849: contig of 4222 bp in length
            *    28850    28949: gap of unknown length
            *    28950    31728: contig of 2779 bp in length
            *    31729    31828: gap of unknown length
            *    31829    35292: contig of 3464 bp in length
            *    35293    35392: gap of unknown length
            *    35393    38559: contig of 3167 bp in length
            *    38560    38659: gap of unknown length
            *    38660    40923: contig of 2264 bp in length
            *    40924    41023: gap of unknown length
            *    41024    43868: contig of 2845 bp in length
            *    43869    43968: gap of unknown length
            *    43969    46193: contig of 2225 bp in length
            *    46194    46293: gap of unknown length
            *    46294    48554: contig of 2261 bp in length
            *    48555    48654: gap of unknown length
            *    48655    51825: contig of 3171 bp in length
            *    51826    51925: gap of unknown length
            *    51926    54521: contig of 2596 bp in length
            *    54522    54621: gap of unknown length
            *    54622    56303: contig of 1682 bp in length
            *    56304    56403: gap of unknown length
            *    56404    58361: contig of 1958 bp in length
            *    58362    58461: gap of unknown length
            *    58462    61315: contig of 2854 bp in length
            *    61316    61415: gap of unknown length
            *    61416    63106: contig of 1691 bp in length
            *    63107    63206: gap of unknown length
            *    63207    65642: contig of 2436 bp in length
            *    65643    65742: gap of unknown length
            *    65743    67378: contig of 1636 bp in length
            *    67379    67478: gap of unknown length
            *    67479    69238: contig of 1760 bp in length
            *    69239    69338: gap of unknown length
            *    69339    71218: contig of 1880 bp in length
            *    71219    71318: gap of unknown length
            *    71319    73232: contig of 1914 bp in length
            *    73233    73332: gap of unknown length
            *    73333    75396: contig of 2064 bp in length
            *    75397    75496: gap of unknown length
            *    75497    77298: contig of 1802 bp in length
            *    77299    77398: gap of unknown length
            *    77399    79999: contig of 2601 bp in length
            *    80000    80099: gap of unknown length
            *    80100    81802: contig of 1703 bp in length
            *    81803    81902: gap of unknown length
            *    81903    83643: contig of 1741 bp in length
            *    83644    83743: gap of unknown length
            *    83744    85906: contig of 2163 bp in length
            *    85907    86006: gap of unknown length
            *    86007    87996: contig of 1990 bp in length
            *    87997    88096: gap of unknown length
            *    88097    89281: contig of 1185 bp in length
            *    89282    89381: gap of unknown length
            *    89382    91687: contig of 2306 bp in length
            *    91688    91787: gap of unknown length
            *    91788    93076: contig of 1289 bp in length
            *    93077    93176: gap of unknown length
            *    93177    95710: contig of 2534 bp in length
            *    95711    95810: gap of unknown length
            *    95811    98026: contig of 2216 bp in length
            *    98027    98126: gap of unknown length
            *    98127    99128: contig of 1002 bp in length
            *    99129    99228: gap of unknown length
            *    99229   101460: contig of 2232 bp in length
            *   101461   101560: gap of unknown length
            *   101561   103011: contig of 1451 bp in length
            *   103012   103111: gap of unknown length
            *   103112   105375: contig of 2264 bp in length
            *   105376   105475: gap of unknown length
            *   105476   107239: contig of 1764 bp in length
            *   107240   107339: gap of unknown length
            *   107340   108355: contig of 1016 bp in length
            *   108356   108455: gap of unknown length
            *   108456   110027: contig of 1572 bp in length
            *   110028   110127: gap of unknown length
            *   110128   111559: contig of 1432 bp in length
            *   111560   111659: gap of unknown length
            *   111660   113741: contig of 2082 bp in length
            *   113742   113841: gap of unknown length
            *   113842   115728: contig of 1887 bp in length
            *   115729   115828: gap of unknown length
            *   115829   116829: contig of 1001 bp in length
            *   116830   116929: gap of unknown length
            *   116930   118826: contig of 1897 bp in length
            *   118827   118926: gap of unknown length
            *   118927   120611: contig of 1685 bp in length
            *   120612   120711: gap of unknown length
            *   120712   121981: contig of 1270 bp in length
            *   121982   122081: gap of unknown length
            *   122082   123463: contig of 1382 bp in length
            *   123464   123563: gap of unknown length
            *   123564   124684: contig of 1121 bp in length
            *   124685   124784: gap of unknown length
            *   124785   125849: contig of 1065 bp in length
            *   125850   125949: gap of unknown length
```

```
          *   125950   126974: contig of 1025 bp in length
          *   126975   127074: gap of unknown length
          *   127075   129000: contig of 1926 bp in length
          *   129001   129100: gap of unknown length

  Query Match             7.7%;  Score 43;  DB 2;  Length 158254;
  Best Local Similarity   60.9%;  Pred. No. 0.54;
  Matches   70;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     316 ctattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgttaacacctct 375
           ||  | |||  | ||| | ||||| |   | | || | | || | || || |||  |  |
Db    4289 CTCGTCTCACTCTGTCATTTAACAAAGGAAACTTGCATACACTCATGGGTGAACCTCAAT 4348

Qy     376 tgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttattaca 430
             | |||| | ||||||| |||||  |  |    |||||||| ||||||
Db    4349 AAATGCTCTGTAACCGTTTCTGGTTATGGAAGGAAACTAGGAAATGTCAATTACA 4403


RESULT  9
AF429315/c
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN


  Query Match             7.4%;  Score 41;  DB 9;  Length 125020;
  Best Local Similarity   12.8%;  Pred. No. 1.9;
  Matches   47;  Conservative  154;  Mismatches  165;  Indels    0;  Gaps    0;

Qy     130 gcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaa 189
           |: ::|  |:|::: :: ::: ||::    | | : :: :: ::  :  ::: :: :::|
Db   66546 GSSWKAMTCYAWMWMWWWWRRWAARWAAGGAAGGMGRWWRRTYWRYWRRRWKGRRSRYWA 66487

Qy     190 tttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctaccac 249
            : |:: : :|: :| :: : :::    | ||::   :|:: ::   :   : : :
Db   66486 MYAGYWGRRRGWYRCAYMYWWRSRKATGKAMGAWYGMWYTRWWKKYKTKWTKKTKGYTCW 66427

Qy     250 gttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattattgat 309
           |:  : : | |  :    : :|::: :  :  :: : : :   :::::  |:: :: :
Db   66426 GYMMMKRMAATWWWTTTWYYYAMRWKYTKKAWYKARWWCMKMCMWWMMWMKAKWSWWAWC 66367

Qy     310 ccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgttaac 369
           :  ::|:        |:  :: : :: :: :  :   : | :|:|:::    : :  :
Db   66366 YGGYMASMTNCASKSASKYRKGYSKKYWMKSYSRGSWSTTYRARARRMMAARRWRRRWGR 66307

Qy     370 acctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttattac 429
           | : :::|::     |::: :    :|    : :: || :  |: :::  :  ::  :||
Db   66306 ATSRSKKGMKWKMAGGMRWTMAARAMCAGSSWRRMSAASRWAGSKRRWSSYTSYWYCYAC 66247

Qy     430 agaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgtactg 489
           | ||    ::  || |:   | :: : || : :|: :  :: ::  |::: ::: :   :
Db   66246 AMAAAMAWWMAAAAATYRGYCWKSMRMGAWRGMAMRMAMMWRWRAKCMMRGYKWCYYRGS 66187

Qy     490 actaaa 495
           |: :::
Db   66186 ASRWWR 66181


RESULT  10
AC094081/c
LOCUS       AC094081              148000 bp    DNA     linear   HTG 14-SEP-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-205D2, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC094081
VERSION     AC094081.1  GI:15617642
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148000)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148000)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 263973
            Center clone name: CIT-HSPC_205D2
            -------
            Summary Statistics
            Consensus quality: 109929 bases at least Q40
            Consensus quality: 130319 bases at least Q30
            Consensus quality: 134832 bases at least Q20
            Estimated insert size: 160000; pulse field gel estimation
            Estimated insert size: 143600; sum-of-contigs estimation
            Quality coverage:  2.79 in Q20 bases; pulse field gel estimation
            Quality coverage:  3.11 in Q20 bases; sum-of-contigs estimation.
```

```
* NOTE: This record contains 45 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1     1345: contig of 1345 bp in length
*     1346     1445: gap of unknown length
*     1446     2511: contig of 1066 bp in length
*     2512     2611: gap of unknown length
*     2612     3875: contig of 1264 bp in length
*     3876     3975: gap of unknown length
*     3976     5634: contig of 1659 bp in length
*     5635     5734: gap of unknown length
*     5735     6916: contig of 1182 bp in length
*     6917     7016: gap of unknown length
*     7017     8138: contig of 1122 bp in length
*     8139     8238: gap of unknown length
*     8239     9444: contig of 1206 bp in length
*     9445     9544: gap of unknown length
*     9545    10707: contig of 1163 bp in length
*    10708    10807: gap of unknown length
*    10808    11934: contig of 1127 bp in length
*    11935    12034: gap of unknown length
*    12035    13169: contig of 1135 bp in length
*    13170    13269: gap of unknown length
*    13270    14619: contig of 1350 bp in length
*    14620    14719: gap of unknown length
*    14720    16497: contig of 1778 bp in length
*    16498    16597: gap of unknown length
*    16598    18234: contig of 1637 bp in length
*    18235    18334: gap of unknown length
*    18335    20429: contig of 2095 bp in length
*    20430    20529: gap of unknown length
*    20530    22800: contig of 2271 bp in length
*    22801    22900: gap of unknown length
*    22901    25553: contig of 2653 bp in length
*    25554    25653: gap of unknown length
*    25654    27672: contig of 2019 bp in length
*    27673    27772: gap of unknown length
*    27773    29756: contig of 1984 bp in length
*    29757    29856: gap of unknown length
*    29857    31386: contig of 1530 bp in length
*    31387    31486: gap of unknown length
*    31487    33886: contig of 2400 bp in length
*    33887    33986: gap of unknown length
*    33987    35894: contig of 1908 bp in length
*    35895    35994: gap of unknown length
*    35995    38807: contig of 2813 bp in length
*    38808    38907: gap of unknown length
*    38908    41243: contig of 2336 bp in length
*    41244    41343: gap of unknown length
*    41344    43289: contig of 1946 bp in length
*    43290    43389: gap of unknown length
*    43390    45626: contig of 2237 bp in length
*    45627    45726: gap of unknown length
*    45727    47930: contig of 2204 bp in length
*    47931    48030: gap of unknown length
*    48031    51418: contig of 3388 bp in length
*    51419    51518: gap of unknown length
*    51519    55433: contig of 3915 bp in length
*    55434    55533: gap of unknown length
*    55534    58444: contig of 2911 bp in length
*    58445    58544: gap of unknown length
*    58545    62187: contig of 3643 bp in length
*    62188    62287: gap of unknown length
*    62288    65728: contig of 3441 bp in length
*    65729    65828: gap of unknown length
*    65829    68678: contig of 2850 bp in length
*    68679    68778: gap of unknown length
*    68779    71850: contig of 3072 bp in length
*    71851    71950: gap of unknown length
*    71951    75194: contig of 3244 bp in length
*    75195    75294: gap of unknown length
*    75295    78698: contig of 3404 bp in length
*    78699    78798: gap of unknown length
*    78799    82388: contig of 3590 bp in length
*    82389    82488: gap of unknown length
*    82489    86689: contig of 4201 bp in length
*    86690    86789: gap of unknown length
*    86790    92232: contig of 5443 bp in length
*    92233    92332: gap of unknown length
*    92333    97989: contig of 5657 bp in length
*    97990    98089: gap of unknown length
*    98090   103184: contig of 5095 bp in length
*   103185   103284: gap of unknown length
*   103285   109565: contig of 6281 bp in length
*   109566   109665: gap of unknown length
*   109666   119541: contig of 9876 bp in length
*   119542   119641: gap of unknown length
*   119642   126139: contig of 6498 bp in length
*   126140   126239: gap of unknown length
*   126240   132394: contig of 6155 bp in length
*   132395   132494: gap of unknown length
*   132495   148000: contig of 15506 bp in length.
FEATURES             Location/Qualifiers
     source          1. .148000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-205D2"
                     /clone_lib="CalTech human BAC library C"
BASE COUNT    39885 a  32077 c  31075 g  40373 t   4590 others
ORIGIN


  Query Match             7.2%;  Score 39.8;  DB 2;  Length 148000;
  Best Local Similarity   49.3%;  Pred. No. 3.9;
  Matches  104;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy      18 tgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttgcggtacctc 77
           || | |   || | || | ||| ||  || |  |   | ||      ||||     ||||
Db   50593 TGGTTTGTTTAAACTACAATCTATACACAGTTCACATATGACATTTACTTGTTAGGCCTC 50534

Qy      78 tacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgttatgcaagagc 137
           |       |   || ||||    |  |     | ||| | |||  ||||     | ||||
Db   50533 TTTTTGTCCAAAACGATGCGGTTCCATACAGCAGCCTCTGGACACATGTAGCTGATGAGC 50474

Qy     138 ccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttgtata 197
           ||||  ||  | | | ||  ||| ||     ||| | | | | |||  |||  || ||
Db   50473 AGATAAAATATGGCTAGTCCAAATTAAGGTATGCTGTAAGTATAAAATGCAAAGTGGATT 50414

Qy     198 cggaaacctaaaggcatcaacaggaacttgc 228
           |||  |   |  ||  || |  |   |||
Db   50413 TCCAAAACAGTATTCAAAAAAAAAAGAATGC 50383


RESULT  11
AC100728
LOCUS       AC100728              196932 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP24-279H4, WORKING DRAFT SEQUENCE, 29 unordered
            pieces.
ACCESSION   AC100728
VERSION     AC100728.1  GI:17048095
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 196932)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

```
  TITLE     Mus musculus, clone RP24-279H4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196932)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L17435
                Center clone name: 279_H_4
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 185765 bases at least Q40
                Consensus quality: 191126 bases at least Q30
                Consensus quality: 193031 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 194132; sum-of-contigs
                Quality coverage: 6.9 in Q20 bases; agarose-fp
                Quality coverage: 6.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7249: contig of 7249 bp in length
            *     7250 7349: gap of      100 bp
            *     7350     7942: contig of 593 bp in length
            *     7943 8042: gap of      100 bp
            *     8043     9463: contig of 1421 bp in length
            *     9464 9563: gap of      100 bp
            *     9564    10770: contig of 1207 bp in length
            *    10771 10870: gap of      100 bp
            *    10871    11894: contig of 1024 bp in length
            *    11895 11994: gap of      100 bp
            *    11995    13096: contig of 1102 bp in length
            *    13097 13196: gap of      100 bp
            *    13197    14617: contig of 1421 bp in length
            *    14618 14717: gap of      100 bp
            *    14718    14939: contig of 222 bp in length
            *    14940 15039: gap of      100 bp
            *    15040    17365: contig of 2326 bp in length
            *    17366 17465: gap of      100 bp
            *    17466    20632: contig of 3167 bp in length
            *    20633 20732: gap of      100 bp
            *    20733    70675: contig of 49943 bp in length
            *    70676 70775: gap of      100 bp
            *    70776    71963: contig of 1188 bp in length
            *    71964 72063: gap of      100 bp
            *    72064    74943: contig of 2880 bp in length
            *    74944 75043: gap of      100 bp
            *    75044    78141: contig of 3098 bp in length
            *    78142 78241: gap of      100 bp
            *    78242    81595: contig of 3354 bp in length
            *    81596 81695: gap of      100 bp
            *    81696    84991: contig of 3296 bp in length
            *    84992 85091: gap of      100 bp
            *    85092    88215: contig of 3124 bp in length
            *    88216 88315: gap of      100 bp
            *    88316    91555: contig of 3240 bp in length
            *    91556 91655: gap of      100 bp
            *    91656    96337: contig of 4682 bp in length
            *    96338 96437: gap of      100 bp
            *    96438   100411: contig of 3974 bp in length
            *   100412 100511: gap of      100 bp
            *   100512   106473: contig of 5962 bp in length
            *   106474 106573: gap of      100 bp
            *   106574   113018: contig of 6445 bp in length
            *   113019 113118: gap of      100 bp
            *   113119   120051: contig of 6933 bp in length
            *   120052 120151: gap of      100 bp
            *   120152   128451: contig of 8300 bp in length
            *   128452 128551: gap of      100 bp
            *   128552   139502: contig of 10951 bp in length
            *   139503 139602: gap of      100 bp
            *   139603   150289: contig of 10687 bp in length
            *   150290 150389: gap of      100 bp
            *   150390   172565: contig of 22176 bp in length
            *   172566 172665: gap of      100 bp
            *   172666   195587: contig of 22922 bp in length
            *   195588 195687: gap of      100 bp
            *   195688   196932: contig of 1245 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196932
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP24-279H4"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .7249
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    7350. .7942
                     /note="assembly_fragment"
     misc_feature    8043. .9463
                     /note="assembly_fragment"
     misc_feature    9564. .10770
                     /note="assembly_fragment"
     misc_feature    10871. .11894
                     /note="assembly_fragment"
     misc_feature    11995. .13096
                     /note="assembly_fragment"
     misc_feature    13197. .14617
                     /note="assembly_fragment"
     misc_feature    14718. .14939
                     /note="assembly_fragment"
     misc_feature    15040. .17365
                     /note="assembly_fragment"
     misc_feature    17466. .20632
                     /note="assembly_fragment"
     misc_feature    20733. .70675
                     /note="assembly_fragment"
```

```
     misc_feature    70776. .71963
                     /note="assembly_fragment"
     misc_feature    72064. .74943
                     /note="assembly_fragment"
     misc_feature    75044. .78141
                     /note="assembly_fragment"
     misc_feature    78242. .81595
                     /note="assembly_fragment"
     misc_feature    81696. .84991
                     /note="assembly_fragment"
     misc_feature    85092. .88215
                     /note="assembly_fragment"
     misc_feature    88316. .91555
                     /note="assembly_fragment"
     misc_feature    91656. .96337
                     /note="assembly_fragment"
     misc_feature    96438. .100411
                     /note="assembly_fragment"
     misc_feature    100512. .106473
                     /note="assembly_fragment"
     misc_feature    106574. .113018
                     /note="assembly_fragment"
     misc_feature    113119. .120051
                     /note="assembly_fragment"
     misc_feature    120152. .128451
                     /note="assembly_fragment"
     misc_feature    128552. .139502
                     /note="assembly_fragment"
     misc_feature    139603. .150289
                     /note="assembly_fragment"
     misc_feature    150390. .172565
                     /note="assembly_fragment"
     misc_feature    172666. .195587
                     /note="assembly_fragment"
     misc_feature    195688. .196932
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    52753 a  45059 c  43133 g  53177 t   2810 others
ORIGIN


  Query Match             6.9%;  Score 38.4;  DB 2;  Length 196932;
  Best Local Similarity   56.2%;  Pred. No. 9.1;
  Matches   72;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy    113 ctgtagacggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggct 172
          |||||||| | |   ||| |    |  |||| |  | | || ||||||  || |||   |
Db  69944 CTGTAGACAGCTACAATGTACTTACATATAATAATAAATAACTCTTAAAAAATAACTCTT 70003

Qy    173 atgactgtgaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtg 232
          || | |    |||  |||   || ||  | |  ||||   || || ||     ||| | |
Db  70004 ATAAATAAATAAATAAATAAATAAACAAACAAATAAATAAATAAAAAGTTTGCTGCCGGG 70063

Qy    233 tggcgtgg 240
            |  ||||
Db  70064 CGTGGTGG 70071


RESULT  12
CEC15A7/c
LOCUS       CEC15A7                 27068 bp    DNA     linear   INV 24-JAN-2002
DEFINITION  Caenorhabditis elegans cosmid C15A7, complete sequence.
ACCESSION   Z67735
VERSION     Z67735.1  GI:1054683
KEYWORDS    HTG; Acetylcholine receptor.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (sites)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    The C.elegans Sequencing Consortium.
REFERENCE   2  (bases 1 to 27068)
  AUTHORS   McMurray,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-1995) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is not the entire insert of clone C15A7.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone C15A7 is at 1 in this sequence. The true
            right end of clone C15A7 is at 16201 in
            sequence Z67738.
            The true left end of clone W03G11 is at 26969 in this sequence. The
            start of this sequence (1. .101) overlaps with the end of sequence
            Z67734.
            The end of this sequence (26969. .27068) overlaps with the start of
            sequence Z67738.
            For a graphical representation of this sequence and its analysis
            see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=C15A7
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
FEATURES             Location/Qualifiers
     source          1. .27068
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="X"
                     /clone="C15A7"
     gene            join(14093. .14194,14238. .14387,14874. .15193,15238. .15393,
                     15524. .15631,15753. .15889,15935. .16147,17039. .17120,
                     17265. .17422,17639. .17855,17902. .18090,18369. .18524,
                     18569. .18673,18943. .19079,19126. .19340)
                     /gene="C15A7.1"
     CDS             join(14093. .14194,14238. .14387,14874. .15193,15238. .15393,
                     15524. .15631,15753. .15889,15935. .16147,17039. .17120,
                     17265. .17422,17639. .17855,17902. .18090,18369. .18524,
                     18569. .18673,18943. .19079,19126. .19340)
                     /gene="C15A7.1"
                     /note="contains similarity to Pfam domain: PF00065
                     (Neurotransmitter-gated ion-channel), Score=15.1,
                     E-value=0.00034, N=2"
                     /codon_start=1
                     /protein_id="CAA91530.1"
                     /db_xref="GI:3874299"
                     /db_xref="SPTREMBL:Q17999"
                     /translation="MIFPLMILVLFNTIPIRVSEADEESERHQAYLDHRKLKEDVLGP
                     LDGSMPSELPVFIGAQLKLDHISRVDEKTGFLTLALTVKFMWSDERLQWDPSHYSGFT
                     RTRLEFVDFWLYQIWKPRIYMTNTYERRQSKTIDLLSNTLVEMDIHSKGFVMTTTKIL
                     LKTFCFLDFKGYPHDFQNCSFSFIPNMNANDIFLGSSMLKAKFTKDFTDPQQVVRVQD
                     FQINDVVTEDLYLFMNRIVLHDLTREPRSLVRSFVSFTIVLQRKNVYVHLQLYAPMIF
                     ICTCLIISGFLPSEYALPLLVFNLISEYLYFHNIKDVLPTDFDGTPTIALIATFTLAE
                     TMSIIGWKMFIITIMTKKQKAYFLNSSKFPLDSGFFNIRVLRRLLTIDKFLMYMLIAQ
                     SFGKPTTIFLQLFFPQRAELRMLTLKVNRQNFPPVLASQYSVKINISFSLEESSYNVS
                     YFHCARYPYSSCGVTMGNGKKLHGFDDDHGFMSVVINAELVSYLVYQPVESFIFQTWI
                     DERLKWNPANYSGVREIVEKTFEFRKDNNCWMPIVKYRSYDRRYSELDLLEFSDARTL
                     ISYKGEIKTALQTMVTTKCQFSFGEYPNDYQNCSIMLIPNQNADEFRFVSPNGYCNPK
                     FENLEHRAVRVHDLHLMGVESNIFYTFANTYFTAEEVTGFEYMAKTEFRFNLMFKRVN
```

```
                     KLLNVKLSIPSILISMFLIIAGLFPNGYSIFGYAYCFFVEVMHGLYVSKILPNDIRGI
                     PYHGALALCFLIETVFLFCWKVFSVYARNQKLFAQLPNVFADNRVFVLKFVFYVDRVV
                     CVALCLQFLSVIYKTR"
     gene            complement(join(19666. .19818,19901. .20144,20196. .20368,
                     20619. .20696,20748. .20829,20875. .20975))
                     /gene="C15A7.2"
     CDS             complement(join(19666. .19818,19901. .20144,20196. .20368,
                     20619. .20696,20748. .20829,20875. .20975))
                     /gene="C15A7.2"
                     /codon_start=1
                     /protein_id="CAA91531.1"
                     /db_xref="GI:3874300"
                     /db_xref="SPTREMBL:Q18000"
                     /translation="MNTIGLGFLVTNYIIYAIDGIGLPIILTSGQFIRALADMLFMYM
                     CLVLARGLNVTKMLLSWLDKLFLVGMFIVFICSYLLMQFWEIKFFDPAEVYAQSESLP
                     AYLLVIWRLTAWVFFVVAALMSRSDSPQKSAFFTSFVFLMTPWFWAPPIFTMIANFCL
                     NNWVRAEVVNIVDNIVTFYGYIIFLVPRKLKHTLKYCVSFSIYRVRQITIKISHSISE
                     LHRLMLMSAENNGNQTNDGMHGVIELTVTNMNTPEKDSDSPAHSDDRGVLERPQIASS
                     "
     gene            complement(join(22061. .22144,22192. .22266,22518. .22628))
                     /gene="C15A7.3"
     CDS             complement(join(22061. .22144,22192. .22266,22518. .22628))
                     /gene="C15A7.3"
                     /codon_start=1
                     /protein_id="CAA91532.1"
                     /db_xref="GI:3874301"
                     /db_xref="SPTREMBL:Q18001"
                     /translation="MKFIQEKHRLVFLLTTFCYHIYIAQGKRAQGILSSKRDFVYLDR
                     FCFQSDTGALEYTFRYPMFYPTQMLLLYFDTEDQWPRAYKELNVG"
     gene            join(complement(Z67738.1:2021. .2043),
                     complement(Z67738.1:631. .1362),
                     complement(Z67738.1:101. .337),complement(26870. .27068),
                     complement(26491. .26655),complement(25596. .25694))
                     /gene="W03G11.4"
     CDS             join(complement(Z67738.1:2021. .2043),
                     complement(Z67738.1:631. .1362),
                     complement(Z67738.1:101. .337),complement(26870. .27068),
                     complement(26491. .26655),complement(25596. .25694))
                     /gene="W03G11.4"
                     /note="cDNA EST yk197a8.5 comes from this gene
                     cDNA EST yk240d1.5 comes from this gene
                     cDNA EST yk466f2.5 comes from this gene
                     cDNA EST yk499c12.3 comes from this gene
                     cDNA EST yk597a3.3 comes from this gene
                     cDNA EST yk240d1.3 comes from this gene
                     cDNA EST yk267f9.3 comes from this gene
                     cDNA EST yk499c12.5 comes from this gene
                     cDNA EST yk503b2.3 comes from this gene
                     cDNA EST yk503b2.5 comes from this gene
                     cDNA EST yk517d6.5 comes from this gene
                     cDNA EST yk597a3.5 comes from this gene"
                     /codon_start=1
                     /protein_id="CAA91533.1"
                     /db_xref="GI:3874302"
                     /db_xref="SPTREMBL:Q18002"
                     /translation="MLDYKGILTNITGFAKTSFYVTLRRGSRNSTNTSVTSTSSASSS
                     KQTEHERLSRDRELVRRHGRLSSSEEARNTAVMGREAREDGIYTEENGAGSPFVVVVS
                     CYNLGEMVVEKSVELENKKMEEGQIPKXKRSRELNIPQRNMFAPDFVIRPDEDLPGFS
                     PIDSEYDTDDDAFEAEELNHRKAPRHTQSMPPPLSGMNPDIPNEYDPVIGIIAIQGVR
                     IPRSRPIDIVPNREPLFPFFGVNSDLQVMLITERMYDVFKERIHSKGKERMLTMMENL
                     AMTLTCGQMCGRLVPVEMMDQTVYYYDYDTVSFLSREELINLLMLCEPDQIEDIMVAS
                     SFPSKFHRETPEQMLQRNKHYQHLMHRPQIFLRRTDYMSCSIDDEDFTFPYFESVDNE
                     HNKRLTLERQLNDDNVDHEEMGALIRTISQDLEEAERQNGLEFVHRSTLALYDARVII
                     PPLCCSSYYSHLSISISININILTFFYVVLESLDEI"
BASE COUNT     9232 a   4367 c   4960 g   8509 t
ORIGIN


  Query Match             6.8%;  Score 38;  DB 3;  Length 27068;
  Best Local Similarity   48.2%;  Pred. No. 12;
  Matches  107;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy    139 cataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttgtatac 198
          ||| |   |   ||  |||||| ||||| |   ||   |||  |||| || |
Db  15755 CATCTGTAAATGAAATGCTTAATAAACAAAGCTAATCCTTGTTTAAAAAAAATAAGGGGA 15696

Qy    199 ggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctaccacgttgcaata 258
            || |||| | | ||    || || | |    || |  ||   | |   |  ||
Db  15695 AATTACTTAAAAACCTAAAATAAAAGTTACAGCCAAGCTTCTAGACTCAAATATTAAAAC 15636

Qy    259 ttggtaagctataaaaatgcttccggagtaacggaaaaaagaattattgatccttcacta 318
          ||        || |||||  ||   | |||| |   || |  |  ||  |   ||  |
Db  15635 TTACGTGCAAATGAAAATCATTGGAGCGTAAAGCTGAAGATGAACATAAACGTTTTTCCT 15576

Qy    319 ttttcaagcggtcctgtaacagatacagcatggagaaacgct 360
          ||    | |  |  || ||| || | || || |   || |||
Db  15575 TTGAAGAACTATTGTGAAACTGACAAAGGATCGCACAAGGCT 15534


RESULT  13
AC092630/c
LOCUS       AC092630              185947 bp    DNA     linear   PRI 08-FEB-2002
DEFINITION  Homo sapiens BAC clone RP11-321C18 from 2, complete sequence.
ACCESSION   AC092630 AC032022
VERSION     AC092630.3  GI:18497222
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185947)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 185947)
  AUTHORS   Pearman,C., Abbott,A., Dixon,R. and Spalding,L.
  TITLE     The sequence of Homo sapiens BAC clone RP11-321C18
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 185947)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 185947)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 185947)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:15624956.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0321C18
                Drafting Center: WIBR
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
```

as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-564H1, 2000 bp overlap; the clone sequenced to the right is RP11-466M21, 2000 bp overlap. Actual start of this clone is at base position 173851 of RP11-564H1; actual end is at base position 9343 of RP11-466M21.

Data from AC012366 was used to finish this clone, AC092630. A single subclone region exists between 47721 and 47759.

The sequence of AC032022 has been incorporated into AC092630.

FEATURES Location/Qualifiers

```
     source          1. .185947
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-321C18"
                     /clone_lib="RPCI-11"
     misc_feature    1. .318
                     /note="similar to EST AW889603 (NID:g8053808)"
     repeat_region   367. .577
                     /rpt_family="Alu"
     repeat_region   723. .1225
                     /rpt_family="MaLR"
     repeat_region   1332. .1397
                     /rpt_family="L2"
     repeat_region   1398. .1699
                     /rpt_family="Alu"
     repeat_region   1700. .1829
                     /rpt_family="L2"
     repeat_region   2028. .2067
                     /rpt_family="AT_rich"
     repeat_region   2364. .2487
                     /rpt_family="L1"
     repeat_region   2483. .2720
                     /rpt_family="L1"
     repeat_region   2714. .2734
                     /rpt_family="AT_rich"
     repeat_region   2738. .3021
                     /rpt_family="Alu"
     repeat_region   2886. .2913
                     /rpt_family="(T)n"
     misc_feature    2914. .3464
                     /note="similar to EST AI808066 (NID:g5394554) wf92c04.x1"
     misc_feature    2914. .3209
                     /note="similar to EST AI470488 (NID:g4332578) tj44e05.x1"
     misc_feature    3003. .3243
                     /note="similar to EST BF874556 (NID:g12264686)"
     misc_feature    3009. .3848
                     /note="similar to EST BG620144 (NID:g13671505)"
     repeat_region   3031. .3056
                     /rpt_family="AT_rich"
     misc_feature    3113. .3650
                     /note="similar to EST AW945629 (NID:g8123384)"
     misc_feature    3506. .3775
                     /note="similar to EST BE004742 (NID:g8264975)"
     misc_feature    3507. .3994
                     /note="similar to EST BG572978 (NID:g13580631)"
     misc_feature    3905. .4229
                     /note="similar to EST R81876 (NID:g858479) yj05h12.r1"
     repeat_region   4041. .4106
                     /rpt_family="Alu"
     repeat_region   5466. .5998
                     /rpt_family="MER2_type"
     repeat_region   6244. .6811
                     /rpt_family="MaLR"
     misc_feature    7420. .7879
                     /note="similar to EST AI921626 (NID:g5657590) wo27h06.x1"
     repeat_region   7421. .7628
                     /rpt_family="L1"
     repeat_region   7834. .8029
                     /rpt_family="MaLR"
     repeat_region   8975. .9220
                     /rpt_family="Alu"
     repeat_region   9416. .9487
                     /rpt_family="(TA)n"
     repeat_region   9976. .10321
                     /rpt_family="MaLR"
     misc_feature    10344. .10772
                     /note="similar to EST BF511583 (NID:g11594881)"
     repeat_region   10350. .10385
                     /rpt_family="L2"
     repeat_region   10641. .10757
                     /rpt_family="MIR"
     misc_feature    12091. .12296
                     /note="similar to EST AV748457 (NID:g10906305)"
     misc_feature    12149. .12296
                     /note="similar to EST BF993178 (NID:g12399501)"
     misc_feature    12171. .12296
                     /note="similar to EST BG232143 (NID:g12727307)"
     repeat_region   12565. .12864
                     /rpt_family="Alu"
     repeat_region   12982. .13123
                     /rpt_family="L1"
     repeat_region   13266. .13287
                     /rpt_family="AT_rich"
     repeat_region   13441. .13683
                     /rpt_family="L1"
     repeat_region   13684. .14454
                     /rpt_family="L1"
     repeat_region   14177. .14263
                     /rpt_family="A-rich"
     repeat_region   14455. .14767
                     /rpt_family="Alu"
     repeat_region   14740. .14762
                     /rpt_family="AT_rich"
     repeat_region   14768. .14869
                     /rpt_family="L1"
     repeat_region   15036. .15327
                     /rpt_family="Alu"
     repeat_region   15354. .15390
                     /rpt_family="(A)n"
     repeat_region   15795. .15985
                     /rpt_family="U2"
     repeat_region   16447. .16470
                     /rpt_family="(TTTTG)n"
     repeat_region   17751. .19325
                     /rpt_family="L1"
     repeat_region   19329. .19387
                     /rpt_family="L1"
     repeat_region   19388. .19679
                     /rpt_family="Alu"
     repeat_region   19680. .20376
                     /rpt_family="L1"
```

```
  Query Match             6.8%;  Score 38;  DB 9;  Length 185947;
  Best Local Similarity   50.5%;  Pred. No. 12;
  Matches   92;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      14 ttcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttgcggta 73
           || |  |  | ||| | |||  || ||| | |  |   | |  |   |  | || |  |
Db 107979 TTTCAAAACTTGCTTCTTTATGTTTTTTTTCCCTTTGCAAATTGTGCCTTTGTTTCTTTC 107920

Qy      74 cctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgttatgcaa 133
           |  ||       | | ||  || ||| |  |   |  |||||||       || |  ||
Db 107919 CTCCTCTTTGTGCTTGACTTTGTATCTCCCTACCACTTCCTGTAACTTAGAATTTTAAAA 107860

Qy     134 gagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttg 193
           |  |  |     | || | |||| || || |||     || || ||| ||| ||||||
Db 107859 CACACATTGTCCAGTGAAATATCTAAACGAGCAAGAATCATCACAGTGCAAACCAATTTC 107800

Qy     194 ta 195
           ||
Db 107799 TA 107798


RESULT  14
AL590155
LOCUS       AL590155              161729 bp    DNA     linear   VRT 16-NOV-2001
DEFINITION  Zebrafish DNA sequence from clone RP71-79P20, complete sequence.
ACCESSION   AL590155
VERSION     AL590155.8  GI:16973913
KEYWORDS    HTG.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 161729)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 17, 2001 this sequence version replaced gi:16151448.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP71-79P20 is
            from a Zebrafish mixed sex BAC library VECTOR: pTARBAC2
            This sequence is the entire insert of clone RP71-79P20.
FEATURES             Location/Qualifiers
     source          1. .161729
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="RP71-79P20"
                     /clone_lib="RPCI-71"
     misc_feature    8778. .8875
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    23043. .23124
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    35627. .35645
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    69443. .69451
                     /note="Inconsistency in the number of bases in
                     mononucleotide run between subclones."
     misc_feature    73674. .73725
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    77054. .77122
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    94519. .94600
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    123588. .123679
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    123744. .123819
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    129853
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 30bp by restriction digest data."
     misc_feature    138831. .138903
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    141877. .141998
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    150970. .151061
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    156194. .156257
                     /note="Tandem repeat. Inconsistency in the number of
                     copies of the repeat element between subclones."
     misc_feature    158289. .158356
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    52162 a  30558 c  28794 g  50215 t
ORIGIN


  Query Match             6.8%;  Score 37.8;  DB 5;  Length 161729;
  Best Local Similarity   47.0%;  Pred. No. 13;
  Matches  117;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy     114 tgtagacggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggcta 173
           | || |   ||||  ||||| |      |  |  | || |  |  | |  | |  |   |
Db 109600 TTTACAGTTATGTACTGCAAAAATAACCAGTAAAATACCATACCGATTATATACAGTTCA 109659

Qy     174 tgactgtgaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgt 233
           |||||  ||| ||| |   |    || |   ||||  | |  |||| ||||| |||  |
Db 109660 CTACTGTAAAATACACTGCCTTGTGGGTAGTTTAAACACTTTTACAGTAACTTACTGCAT 109719

Qy     234 ggcgtggagctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacgga 293
               | |  | |    ||  | | |    ||   |   || | ||||   ||||  ||
Db 109720 ATTAGGAATTTGCAGTATTCTACTGTAATCAAAGCAGTCAAGTACTTCTACTGTAATTGA 109779

Qy     294 aaaaagaattattgatccttcactattttcaagcggtcctgtaacagatacagcatggag 353
           | |    |  | |      |  |  |||| |      |||| || | |  | |   ||
Db 109780 AGACTACAGATTACAGTAAAAATGAAGTTCATGAAAAAATGTATCATAAAATGTAATTAG 109839

Qy     354 aaacgcttg 362
           ||  | |||
Db 109840 AATAGATTG 109848


RESULT  15
AC022390
LOCUS       AC022390              161221 bp    DNA     linear   PRI 21-AUG-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-131N11, complete sequence.
ACCESSION   AC022390
VERSION     AC022390.8  GI:15217186
```

```
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161221)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161221)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 161221)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Aug 21, 2001 this sequence version replaced gi:15011655.
FEATURES             Location/Qualifiers
     source          1. .161221
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-131N11"
                     /clone_lib="RPCI-11"
BASE COUNT    51066 a  31550 c  30263 g  48342 t
ORIGIN


  Query Match             6.8%;  Score 37.6;  DB 9;  Length 161221;
  Best Local Similarity   61.0%;  Pred. No. 15;
  Matches   61;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     146 tgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacc 205
           ||| | | |  | ||| ||  | |||||| | | ||||| || |||  ||||   |||
Db   77122 TGAAATATAGTTAAATAAATTATGGCTATTAATATGAAACACCATTAATATAAAAAAAAT 77181

Qy     206 taaaggcatcaacaggaacttgctgtgtggcgtggagcta 245
           ||||   | ||||  |||  |   | |  || || | |||
Db   77182 TAAAAAGAGCAACCAGAATATTTGGAGATGCTTGTAACTA 77221


Search completed: June 27, 2002, 20:16:15
Job time: 9703 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 18:35:02 ; Search time 362.26 Seconds
(without alignments)
2630.397 Million cell updates/sec

Title: US-09-727-769A-5
Perfect score: 555
Sequence: 1 ttggcgagtgtaattcctga.....atgtatccagctgtggattt 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	555	100.0	555	22	AAF90280	Nucleotide sequenc
	2	555	100.0	1080	22	AAF90281	Nucleotide sequenc
	3	344.2	62.0	555	21	AAZ49494	Chryseobacterium g
	4	344.2	62.0	1080	21	AAZ49495	Chryseobacterium g
c	5	37	6.7	15865	22	AAK90825	Human digestive sy
c	6	36	6.5	19576	24	AAS61258	Human gene regulat
c	7	35.6	6.4	2489	21	AAZ90461	E. canis 120 kDa i
c	8	35.4	6.4	9300	22	ABA15594	Human nervous syst
	9	35	6.3	4383	23	ABL18719	Drosophila melanog
	10	35	6.3	4590	22	AAH24065	Yeast AOD9604-asso
c	11	35	6.3	106416	23	ABL18718	Drosophila melanog
	12	34.8	6.3	1751	19	AAV40300	Sugarbeet choline
c	13	34.4	6.2	1072	18	AAV74746	Staphylococcus aur
c	14	34.4	6.2	1570	22	ABA18492	Human nervous syst
c	15	34.4	6.2	3943	22	AAS31317	Human cDNA encodin
c	16	34.4	6.2	12582	22	AAK84357	Human immune/haema
c	17	34.4	6.2	23821	21	AAA92469	Shewanella sp. SCR
	18	34.2	6.2	2019	22	AAF25737	Tomato LIN7 promot
	19	34.2	6.2	4197	16	AAQ99430	B. sphaericus SLP
c	20	34	6.1	3581	22	AAD09377	Human vesicle traf
c	21	34	6.1	3945	22	AAS31428	Human cDNA encodin
c	22	33.8	6.1	9289	22	AAS46502	Tumour suppressor
c	23	33.6	6.1	4223	22	AAH78011	Nucleotide sequenc
c	24	33.4	6.0	932	22	AAH29847	S cerevisiae apopt
c	25	33.2	6.0	397	14	AAQ60606	Human brain Expres
	26	33	5.9	3442	21	AAC59824	Human secreted pro
c	27	33	5.9	5195	24	ABL33596	Human immune syste
	28	32.8	5.9	1949	11	AAQ06018	Sequence encoding
	29	32.8	5.9	1950	10	AAN91094	Protein G gene. S
	30	32.8	5.9	1950	11	AAQ06017	Sequence encoding
	31	32.8	5.9	1950	12	AAQ10001	Sequecne encoding
	32	32.8	5.9	1950	15	AAQ64644	Streptococcus Prot
c	33	32.8	5.9	5366	20	AAX58001	Gonadotropic hormo
c	34	32.8	5.9	5366	20	AAX56273	Human apoptosis in
c	35	32.8	5.9	5502	17	AAT30092	Neuronal apoptosis
c	36	32.8	5.9	5502	18	AAT71263	Neuronal apoptosis
c	37	32.8	5.9	5984	20	AAX58000	Gonadotropic hormo
c	38	32.8	5.9	5984	20	AAX56272	Human apoptosis in
c	39	32.8	5.9	5984	21	AAA39808	Human NAIP cDNA.
c	40	32.8	5.9	6124	18	AAT71265	Neuronal apoptosis
c	41	32.8	5.9	6133	18	AAT71264	Neuronal apoptosis
c	42	32.8	5.9	6228	18	AAT71266	Neuronal apoptosis
c	43	32.8	5.9	28473	19	AAV52216	Streptococcus pneu
	44	32.6	5.9	13431	22	AAD08701	Lycopersicon penne
c	45	32.4	5.8	575	22	AAS30852	Human cDNA encodin

ALIGNMENTS

RESULT 1
AAF90280
ID AAF90280 standard; DNA; 555 BP.
XX
AC AAF90280;
XX
DT 22-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a protein-deamidating enzyme.
XX
KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW bakery; confectionery; ss.
XX
OS Cryseobacterium sp.
XX
PN EP1106696-A1.
XX
PD 13-JUN-2001.
XX
PF 04-DEC-2000; 2000EP-0310768.
XX
PR 03-DEC-1999; 99JP-0345044.
XX
PA (AMAN-) AMANO ENZYME INC.
XX
PI Yamaguchi S;
XX
DR WPI; 2001-376907/40.
DR P-PSDB; AAB84386.
XX
PT New enzyme for use in e.g. bakery has an ability to deamidate amido
PT groups in a protein -

```
XX
PS   Claim 8; Page 22; 43pp; English.
XX
CC   The present sequence encodes a protein-deamidating enzyme from
CC   Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
CC   groups in a protein by directly acting upon the amido groups without
CC   cutting peptide bonds and without cross-linking the protein. The enzyme
CC   thus reduces the mineral sensitivity of the protein and increases the
CC   soluble mineral content in the protein-mineral solution, improving the
CC   absorption of minerals in the human body. The enzyme reduces or removes
CC   the toxicity and allergenic property of the protein in a food
CC   e.g. allergy. The enzyme is useful for the improvement of dough in the
CC   field of bakery and confectionery e.g. for the production of crackers,
CC   biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC   protein products, in various food articles e.g. meat or fish products
CC   and noodles; and for improving functionality of plant or animal protein.
XX
SQ   Sequence 555 BP; 166 A; 119 C; 107 G; 163 T; 0 other;


  Query Match             100.0%;  Score 555;  DB 22;  Length 555;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-160;
  Matches  555;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 60

Qy      61 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 120

Qy     121 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 180

Qy     181 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 240

Qy     241 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 300

Qy     301 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 360

Qy     361 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 420

Qy     421 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 480

Qy     481 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 540

Qy     541 tccagctgtggattt 555
           |||||||||||||||
Db     541 tccagctgtggattt 555


RESULT   2
AAF90281
ID   AAF90281 standard; DNA; 1080 BP.
XX
AC   AAF90281;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of a protein-deamidating enzyme.
XX
KW   Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW   bakery; confectionery; ss.
XX
OS   Cryseobacterium sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             61..1023
FT                   /*tag= a
FT                   /product= "protein-deamidating enzyme"
XX
PN   EP1106696-A1.
XX
PD   13-JUN-2001.
XX
PF   04-DEC-2000; 2000EP-0310768.
XX
PR   03-DEC-1999;   99JP-0345044.
XX
PA   (AMAN-) AMANO ENZYME INC.
XX
PI   Yamaguchi S;
XX
DR   WPI; 2001-376907/40.
DR   P-PSDB; AAB84387.
XX
PT   New enzyme for use in e.g. bakery has an ability to deamidate amido
PT   groups in a protein -
XX
PS   Example 11; Page 23; 43pp; English.
XX
CC   The present sequence encodes a protein-deamidating enzyme from
CC   Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
CC   groups in a protein by directly acting upon the amido groups without
CC   cutting peptide bonds and without cross-linking the protein. The enzyme
CC   thus reduces the mineral sensitivity of the protein and increases the
CC   soluble mineral content in the protein-mineral solution, improving the
CC   absorption of minerals in the human body. The enzyme reduces or removes
CC   the toxicity and allergenic property of the protein in a food
CC   e.g. allergy. The enzyme is useful for the improvement of dough in the
CC   field of bakery and confectionery e.g. for the production of crackers,
CC   biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC   protein products, in various food articles e.g. meat or fish products
CC   and noodles; and for improving functionality of plant or animal protein.
XX
SQ   Sequence 1080 BP; 371 A; 204 C; 203 G; 302 T; 0 other;


  Query Match             100.0%;  Score 555;  DB 22;  Length 1080;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-160;
  Matches  555;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     466 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 525

Qy      61 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     526 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 585

Qy     121 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     586 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 645

Qy     181 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     646 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 705

Qy     241 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     706 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 765
```

```
Qy    301 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    766 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 825

Qy    361 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    826 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 885

Qy    421 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    886 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 945

Qy    481 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    946 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 1005

Qy    541 tccagctgtggattt 555
          |||||||||||||||
Db   1006 tccagctgtggattt 1020
```

RESULT 3
AAZ49494
ID AAZ49494 standard; DNA; 555 BP.
XX
AC AAZ49494;
XX
DT 04-APR-2000 (first entry)
XX
DE Chryseobacterium gleum protein-deamidating enzyme encoding DNA.
XX
KW Protein deamidating enzyme; soil bacterium; deamidation activity;
KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW protein engineering; surface hydrophobicity; toxicity; allergic;
KW mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.
XX
OS Chryseobacterium gleum 'JCM 2410'.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..555
FT /*tag= a
FT /product= "Protein deamidating enzyme"
FT /function= "Deamidate amido groups in a protein"
FT /note= "Improves protein function"
XX
PN EP976829-A2.
XX
PD 02-FEB-2000.
XX
PF 04-JUN-1999; 99EP-0304367.
XX
PR 04-JUN-1998; 98JP-0173940.
XX
PA (AMAN) AMANO PHARM KK.
XX
PI Yamaguchi S, Matsuura A;
XX
DR WPI; 2000-118552/11.
DR P-PSDB; AAY44582.
XX
PT New enzyme for modifying and improving the function of proteins and/or
PT peptides has deamidating activity without causing cross linking -
XX
PS Claim 8; Page 23; 57pp; English.
XX
CC The present sequence is the DNA encoding the protein-deamidating enzyme,
CC isolated from a new strain of soil bacterium, Chryseobacterium gleum
CC JCM 2410. The enzyme exerts the deamidation activity by directly acting
CC upon side chain amido groups in the protein in bonded state and
CC releasing side chain carboxyl groups and ammonia. It can deaminate high
CC molecular weight proteins, without cross linking and cleavage of peptide
CC bonds, to improve protein function. This sequence is used for protein
CC engineering, to cause an increase in surface hydrophobicity and improve
CC the function of a plant or animal protein. It can also be used to remove
CC or reduce toxicity or allergic property of proteins in food, decrease
CC mineral sensitivity of protein, to allow greater absorption into the body
CC and to solubilise calcium for use in drinks and mineral enhancing agents.
XX
SQ Sequence 555 BP; 159 A; 122 C; 112 G; 162 T; 0 other;

```
  Query Match             62.0%;  Score 344.2;  DB 21;  Length 555;
  Best Local Similarity   76.7%;  Pred. No. 1.9e-95;
  Matches  421;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy      7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
          ||||| |||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db      7 agtgttattcctgatctggcaacgctgaacagtttatttacccagatcaaaaaccaggct 66

Qy     67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
          ||||| || ||||| || || || || || ||||| |||||||||| || |||||||||
Db     67 tgcggaacttctacagcatcttctccttgtatcaccttcagatatccggttgacggatgt 126

Qy    127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    127 tatgcaagggctcacaaaatgagacaaatcctattgaacgccggctatgactgtgaaaag 186

Qy    187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    187 cagttcgtatatggtaatctgagagcttctacaggaacatgctgtgtatcatgggtatat 246

Qy    247 cacgttgcaatattggtaagctataaaaaatgcttccggagtaacggaaaaaagaattatt 306
          ||||| ||||| ||||||||||  ||||||||||| ||| |    |||||||||| ||
Db    247 cacgtagcaattttggtaagcttcaaaaatgcttcaggaattgttgaaaaaagaatcata 306

Qy    307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
          ||||||||| |||| || |||||||||||||||||| | |||||||||   || ||
Db    307 gatccttcattattctccagcggtcctgtaacagattctgcatggagagctgcatgtacc 366

Qy    367 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
          |||||    |||||||||||| || || || || || || ||||| ||||||||||||||
Db    367 aacacaagctgcggatctgcgtctgtatcttcctacgccaatacagcaggaaatgtttac 426

Qy    427 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 486
          ||||||||||      ||| |  |||||||| |||||     | ||||||| ||||||
Db    427 tacagaagtccgtcaggttcattactgtatgataacaactatgtgaataccaattgtgta 486

Qy    487 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 546
           | |  | ||| |||   ||||| |||||||||||||| || |||||   |||| | |||
Db    487 ttaaacatattctcatccctttcaggatgttctccttccccagcaccaagtgtagcaagc 546

Qy    547 tgtggattt 555
          |||||||||
Db    547 tgtggattt 555
```

RESULT 4
AAZ49495
ID AAZ49495 standard; DNA; 1080 BP.
XX
AC AAZ49495;
XX
DT 04-APR-2000 (first entry)
XX
DE Chryseobacterium gleum protein-deamidating prepro-enzyme encoding gene.
XX
KW Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;
KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW protein engineering; surface hydrophobicity; toxicity; allergic;
KW mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.
XX
OS Chryseobacterium gleum 'JCM 2410'.
XX
FH Key Location/Qualifiers
FT CDS 61..1020
FT /*tag= a

```
FT                   /product= "C. gleum protein deamidating prepro-enzyme"
FT                   /function= "Deamidate amido groups in a protein"
FT   sig_peptide     61..123
FT                   /*tag=  b
FT                   /note= "Corresponds to pre region of the protein"
FT   mat_peptide     463..1017
FT                   /*tag=  c
FT                   /label= Mature_protein_deamidating_enzyme
FT                   /note= "Improves protein function"
XX
PN   EP976829-A2.
XX
PD   02-FEB-2000.
XX
PF   04-JUN-1999;   99EP-0304367.
XX
PR   04-JUN-1998;   98JP-0173940.
XX
PA   (AMAN ) AMANO PHARM KK.
XX
PI   Yamaguchi S,  Matsuura A;
XX
DR   WPI; 2000-118552/11.
DR   P-PSDB; AAY44583.
XX
PT   New enzyme for modifying and improving the function of proteins and/or
PT   peptides has deamidating activity without causing cross linking -
XX
PS   Example 26; Page 25; 57pp; English.
XX
CC   The present sequence is the gene encoding the protein-deamidating
CC   prepro-enzyme, isolated from a new strain of soil bacterium,
CC   Chryseobacterium gleum JCM 2410. The enzyme exerts the deamidation
CC   activity by directly acting upon side chain amido groups in the protein
CC   in bonded state and releasing side chain carboxyl groups and ammonia. It
CC   can deaminate high molecular weight proteins, without cross linking and
CC   cleavage of peptide bonds, to improve protein function. This sequence is
CC   used for protein engineering, to cause an increase in surface
CC   hydrophobicity and improve the function of a plant or animal protein. It
CC   can also be used to remove or reduce toxicity or allergic property of
CC   proteins in food, decrease mineral sensitivity of protein, to allow
CC   greater absorption into the body and to solubilise calcium for use in
CC   drinks and mineral enhancing agents.
XX
SQ   Sequence 1080 BP; 354 A; 210 C; 205 G; 311 T; 0 other;


  Query Match             62.0%;  Score 344.2;  DB 21;  Length 1080;
  Best Local Similarity   76.7%;  Pred. No. 2.5e-95;
  Matches  421;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy      7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
          ||||| |||||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db    469 agtgttattcctgatctggcaacgctgaacagtttatttacccagatcaaaaaccaggct 528

Qy     67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
          ||||| || ||||| || || || || || ||||| |||||||||| || |||||||||
Db    529 tgcggaacttctacagcatcttctccttgtatcaccttcagatatccggttgacggatgt 588

Qy    127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    589 tatgcaagggctcacaaaatgagacaaatcctattgaacgccggctatgactgtgaaaag 648

Qy    187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    649 cagttcgtatatggtaatctgagagcttctacaggaacatgctgtgtatcatgggtatat 708

Qy    247 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 306
          ||||| ||||| |||||||||| |||||||||| ||| |    |||||||||| ||
Db    709 cacgtagcaattttggtaagcttcaaaaatgcttcaggaattgttgaaaaaagaatcata 768

Qy    307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
          ||||||||| |||| || |||||||||||||||||| | |||||||||   || ||

Db    769 gatccttcattattctccagcggtcctgtaacagattctgcatggagagctgcatgtacc 828

Qy    367 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
          |||||    |||||||||| || || || || || || ||||| ||||||||||||||
Db    829 aacacaagctgcggatctgcgtctgtatcttcctacgccaatacagcaggaaatgtttac 888

Qy    427 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 486
          ||||||||||      ||| |  |||||||| |||||     | ||||||| ||||||
Db    889 tacagaagtccgtcaggttcattactgtatgataacaactatgtgaataccaattgtgta 948

Qy    487 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 546
            | |  | ||| |||   ||||| |||||||||||||| || |||||   |||| | |||
Db    949 ttaaacatattctcatccctttcaggatgttctccttccccagcaccaagtgtagcaagc 1008

Qy    547 tgtggattt 555
          |||||||||
Db   1009 tgtggattt 1017


RESULT   5
AAK90825/c
ID   AAK90825 standard; DNA; 15865 BP.
XX
AC   AAK90825;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 4401.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
```

PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 4401; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 15865 BP; 4775 A; 3204 C; 3309 G; 4577 T; 0 other;

Query Match 6.7%; Score 37; DB 22; Length 15865;
Best Local Similarity 50.9%; Pred. No. 1.4;
Matches 88; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy     54 aaagaatcagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcc 113
          || || |   |||   |||||  ||  |   |    | ||| |||  ||   | | |
Db  13604 AATGACTAGCTCTCAAGGTACTCCTTGGTACTTGAAAACATTCATTTCAGAAAAAAAAAA 13545

Qy    114 tgtagacggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggcta 173
            | |   |  |||      ||   ||||  | ||| |||||  | |||   ||| |||
Db  13544 AACAAAAAAACTTTACATGGAAGGAAATAAATTTAGAGAAATCATTATGCCTAACAGCTG 13485

Qy    174 tgactgtgaaaaacaatttgtatacggaaacctaaaggcatcaacaggaactt 226
            || | || |||  ||| ||| |  ||   || ||| |     |   ||||
Db  13484 GTACAGAGAGAAAATATTGGTAAAATTAAGGATATAGGTAAATGTACAGACTT 13432
```

RESULT 6

```
AAS61258/c
ID   AAS61258 standard; DNA; 19576 BP.
XX
AC   AAS61258;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human gene regulation-associated gene oligonucleotide #213.
XX
KW   Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW   cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW   asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW   renal disease; Preeclampsia; cardiac allograft vascular disease;
KW   colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW   immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW   nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200177375-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP03968.
XX
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-017470/02.
XX
PT   New nucleic acid sequences from chemically modified genes associated
PT   with gene regulation, useful for analysing cytosine methylations for
PT   diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT   disease -
XX
PS   Claim 1; SEQ ID No 219; 26pp; English.
XX
CC   The invention relates to 224 nucleic acid sequences comprising at least
CC   18 bases of a chemically pretreated gene associated with gene regulation
CC   selected from 43 known genes (or complementary sequences). The
CC   chemical pretreatment converts cytosine bases unmethylated at the
CC   5-position to uracil or another base with hybridisation behaviour
CC   dissimilar to cytosine, to enable analysis of cytosine methylations.
CC   The DNA sequences, oligomers (or sets/arrays) and method are
CC   useful in the diagnosis of diseases (or predisposition to diseases)
CC   associated with gene regulation and in therapy of such diseases, by
CC   enabling analysis of the cytosine methylation patterns of such genes,
CC   kits are provided. They are especially useful in diagnosis
CC   and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC   disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC   asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC   preeclampsia, graft versus-host disease. The present sequence is a
CC   sequence included in the sequence data for this specification and is
CC   associated with the human gene regulation-associated genes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 19576 BP; 5357 A; 307 C; 4161 G; 9750 T; 1 other;

  Query Match             6.5%;  Score 36;  DB 24;  Length 19576;
  Best Local Similarity   49.5%;  Pred. No. 3;
  Matches   93;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy    312 ttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgttaacac 371
          ||| ||||||| ||    ||   |||| ||     ||  | |  |        || |
Db   6024 TTCTCTATTTTAAATTAATCTACTAACCTATTATAAATAAACATCCATAAATTCTACCCT 5965

Qy    372 ctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttattacag 431
            ||     |  |  | |  |||| ||| |    ||||    |  | || |||||  | |
Db   5964 AACTAAAAAACATTTAACTTTTTCGTCTAAAATAAATATATAAATATATATTTATAAAAA 5905

Qy    432 aagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgtactgac 491
          ||     | |  || | |||   ||   || ||| |  | ||||   |||   ||   ||
Db   5904 AAAAATAACTCCTTTTAACCAAAATATTAATAATTTTCTAAATAAATACTTACTATCCAC 5845

Qy    492 taaatttt 499
            ||| |||
Db   5844 AAAAATTT 5837


RESULT   7
AAZ90461/c
ID   AAZ90461 standard; DNA; 2489 BP.
XX
AC   AAZ90461;
XX
DT   06-JUN-2000  (first entry)
XX
DE   E. canis 120 kDa immunodominant antigenic protein encoding DNA.
XX
KW   120 kDa protein; immunodominant; antigen; immunoreactive; vaccine;
KW   Ehrlichia canis infection; antibacterial; ds.
XX
OS   Ehrlichia canis.
XX
FH   Key             Location/Qualifiers
FT   CDS             341..2407
FT                   /*tag= a
FT                   /product= "120 kDa immunodominant antigenic protein"
XX
PN   WO200012688-A1.
XX
PD   09-MAR-2000.
XX
PF   27-AUG-1999;   99WO-US19538.
XX
PR   27-AUG-1998;   98US-0141047.
XX
PA   (RERE-) RES DEV FOUND.
XX
PI   Walker DH,  Yu X;
XX
DR   WPI; 2000-256636/22.
DR   P-PSDB; AAY57274.
XX
PT   Protein immunoreactive with anti-Ehrlichia canis and comprises a
PT   sequence of 688 amino acids, useful for inhibiting Ehrlichia canis
PT   infection -
XX
PS   Claim 2; Fig 3; 78pp; English.
XX
CC   The invention provides a 120 kDa immunodominant antigenic protein of
CC   Ehrlichia canis that is immunoreactive with anti-Ehrlichia canis
CC   serum. The protein can be expressed by standard recombinant methodology.
CC   The protein is useful for inhibiting Ehrlichia canis infection. The
CC   present sequence represents a DNA encoding the immunodominant protein
CC   of E. canis.
XX
SQ   Sequence 2489 BP; 935 A; 286 C; 552 G; 716 T; 0 other;

  Query Match             6.4%;  Score 35.6;  DB 21;  Length 2489;
  Best Local Similarity   45.9%;  Pred. No. 1.6;
  Matches  122;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy    276 tgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtcctgt 335
          | |||||  | ||     | | |   ||  |    |||||||   |  | | |  | |
```

```
Db   1843 TTCTTCCTTACTAGTTTCAGATACTTTTTCTCCAACTTCACTTGATGAATGTTCTACACT 1784

Qy    336 aacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtttc 395
          | ||  ||||||| |  | ||  ||| || |   || ||  | | || | | ||  |
Db   1783 ACCATCTACAGCAGGTTGCAAATCTTCCGCTTTAACTTCAGGAGTATTTTCCTCTTTACT 1724

Qy    396 ctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctgta 455
             ||  | || |  | |   || |     | |   | || |||      | |
Db   1723 AGTTTCAGATACTTTTTCTCCAACTTCACTTGATGAATGTTCTACACTACCATCTACAGC 1664

Qy    456 tgacaacaatctgatcaataccaactgtgtactgactaaattttcactgctttccggatg 515
           |   |||         |   |  |  | | |   |   | || |||  |||| |
Db   1663 AGGTTGCAAATCTTCTGCTTTAACTTCAGGAGTATTTTCCTCTTTACTAGTTTCAGATAC 1604

Qy    516 ttctccttcacctgcaccggatgtat 541
          || | || || || |||  |||| ||
Db   1603 TTTTTCTCCAACTTCACTTGATGAAT 1578


RESULT   8
ABA15594/c
ID   ABA15594 standard; DNA; 9300 BP.
XX
AC   ABA15594;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 7925.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
```

```
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 7925; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 9300 BP; 2858 A; 1566 C; 1864 G; 3012 T; 0 other;

  Query Match             6.4%;  Score 35.4;  DB 22;  Length 9300;
  Best Local Similarity   49.2%;  Pred. No. 3.3;
  Matches   93;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
Db    5954 AATTTTTTTTTTTTGGTAGAGACAAAGGTCTTGCTATGTTACCCAAGCTGGTCTTGATCT 5895

Qy      67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
Db    5894 CCTGGTGCCAAACAAACCTCCCACCCTGGCCTCCCATTTAAAAATATTTTTGATGCTACT 5835

Qy     127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
Db    5834 CATTCAAGGGCCTTTTAAATAAGGAACATTACAATTTTCATTTTCTTCATTTTTTAAAAA 5775

Qy     187 caatttgta 195
Db    5774 ATATTTTTA 5766


RESULT   9
ABL18719
ID   ABL18719 standard; DNA; 4383 BP.
XX
AC   ABL18719;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 7630.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 7630; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4383 BP; 1563 A; 942 C; 865 G; 1013 T; 0 other;

  Query Match             6.3%;  Score 35;  DB 23;  Length 4383;
  Best Local Similarity   57.9%;  Pred. No. 3.2;
  Matches   62;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;
```

```
Qy      93 atgcatcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagaca 152
           ||||| | || | | | ||  ||| |  | |  |  | ||| |  | |  | |  |   |
Db     654 atgcaacccagtaacaaatgatgttgcagcaactactacaatacacgagtataccaagaa 713

Qy     153 aatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacg 199
           |||   |  |||||| || | | |    || ||||| |||||||||
Db     714 catcaacaacaacaactgcaacggcaaccaacaacaaattgtatacg 760


RESULT 10
AAH24065
ID   AAH24065 standard; DNA; 4590 BP.
XX
AC   AAH24065;
XX
DT   29-AUG-20Q1  (first entry)
XX
DE   Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
XX
KW   Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
KW   modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
KW   lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
KW   functional food; transgenic yeast; fat/lean ratio; food use; ds.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    10
FT                   /*tag=  a
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3617
FT                   /*tag=  b
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3649
FT                   /*tag=  c
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3679
FT                   /*tag=  d
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3819
FT                   /*tag=  e
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3862
FT                   /*tag=  f
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3864
FT                   /*tag=  g
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3888
FT                   /*tag=  h
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3890
FT                   /*tag=  i
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3912
FT                   /*tag=  j
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3914
FT                   /*tag=  k
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3938
FT                   /*tag=  l
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3939
FT                   /*tag=  m
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3941
FT                   /*tag=  o
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3943
FT                   /*tag=  p
FT                   /note= "Represented as * in the specification"
FT   misc_feature    4361
FT                   /*tag=  q
FT                   /note= "Represented as * in the specification"
XX
PN   WO200133977-A1.
XX
PD   17-MAY-2001.
XX
PF   06-NOV-2000; 2000WO-AU01362.
XX
PR   05-NOV-1999;   99AU-0003875.
XX
PA   (META-) METABOLIC PHARM LTD.
XX
PI   Belyea CI,  Ng FM,  Vaughan P;
XX
DR   WPI; 2001-328876/34.
XX
PT   New organisms containing nucleic acid encoding a growth hormone
PT   fragment which modulates lipid metabolism are useful to produce dietary
PT   aids for obesity and in the meat production industry  -
XX
PS   Disclosure; Page 48-50; 54pp; English.
XX
CC   The invention relates to novel transgenic organisms useful in the
CC   production of functional food and drink products for the treatment
CC   or prevention of obesity via the regulation of lipid metabolism. The
CC   organisms comprise a polynucleotide encoding a growth hormone fragment
CC   capable of stimulating the activity of hormone-sensitive lipase (the key
CC   enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
CC   enzyme in lipogenesis). The growth hormone fragment preferably contains
CC   at least the disulphide-bonded loop of a mammalian growth hormone (but is
CC   not the full-length growth hormone) and is optionally linked to an
CC   epitope tag or heterologous fusion protein partner. The transgenic
CC   organism may be a microorganism used to produce a fermented product
CC   (e.g., yeast), or an edible plant or animal or cell thereof. Food or
CC   drink made using methods of the invention are used to modify fat/lean
CC   ratio, lipid metabolism or food use in a mammal. In particular, the food
CC   or drink products may be used to treat or prevent obesity, particularly
CC   in humans, and may also be used to improve the fat/lean ration of
CC   livestock raised for meat production. In the exemplification of the
CC   invention, the human growth hormone (hGH) fragment analogue AOD9604 was
CC   expressed in yeast, optionally fused to the FLAG epitope (AAB73625).
CC   The present sequence is described as a DNA sequence from yeast in
CC   the sequence listing, but is not further referred to in the
CC   specification.
XX
SQ   Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;


  Query Match             6.3%;  Score 35;  DB 22;  Length 4590;
  Best Local Similarity   11.2%;  Pred. No. 3.2;
  Matches   53;  Conservative  228;  Mismatches  192;  Indels    2;  Gaps    1;

Qy      34 aattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctcacca 93
           .::::|: :  : ::: :::::::::: :  : : :::::: : :::  :   : :::
Db    3880 rvhdcdvwbnhnvwddhdhdhdhdhdhdnwchnhnbrvhdhdhdddvyymnrghmhmhnn 3939

Qy      94 tgcatcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagacaa 153
           : : ||:::::::  | : ::|::::::::  :  : :::: :::: ::: |:  :  | :
Db    3940 hnhntcdhdhdhdhntchabdgkmvvvvvvvvvvvvvvvvvchkhhdhdhkghmcmhcccm 3999

Qy     154 atcttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaaggca 213
           :::::||   ::| :  ::|:  |:  :| |:| :|: |  : : ::|  ::   ::::
Db    4000 hkmdhaaaanrmacmhhhhadngckhadataragrahnthadrrtrragnmbrnnmbrs- 4058

Qy     214 tcaacaggaacttgctgtgtggcgtggagctaccacgttgcaatattggtaagctataaa 273
            |::: : :::::   :::  :: ||: : :   ::: ::::: :: :: : ::  |:|
Db    4059 -cdmmckmmmmkwykndbdvwryytgbbdrknwndysmbshmwndyskdcwndysctmat 4117

Qy     274 aatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtcct 333
           :  ::  : ::: :  |  | : ::  |   :: :::||     : :  :| :  :::||
Db    4118 rcvrysavasdwndysctmatrcvrysavasdwndysctmatrcvrysavasdwndysct 4177
```

```
Qy    334 gtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtt 393
          :|::  |:|: : :    :| :|   :    :   : ||   : : :  :|   : :
Db   4178 matrcvrysavasdwndysctmatrcvrysavasdwndysctmatrcvrysavasdrksk 4237

Qy    394 tcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctg 453
          : :: :   |    :::  :  :|:| :  :::: :: : :   ::::  |  |  | |
Db   4238 wwskkwssmgtmsnwrmansymbaracrrnwhhbbsdhmbsstdwrddcstmatsscdtw 4297

Qy    454 tatgacaacaatctgatcaataccaactgtgtactgactaaattttcactgcttt 508
          :::: : ::: :: ::: : ::|:  :::  ::| :  : :  |::||: :||:
Db   4298 ndyswndyshhsswndysndndscwndysdmcrstwrdrrsdgrthhackchttd 4352


RESULT  11
ABL18718/c
ID   ABL18718 standard; DNA; 106416 BP.
XX
AC   ABL18718;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 7627.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 7627; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 106416 BP; 30148 A; 22327 C; 22141 G; 31800 T; 0 other;


  Query Match             6.3%;  Score 35;  DB 23;  Length 106416;
  Best Local Similarity   57.9%;  Pred. No. 13;
  Matches   62;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     93 atgcatcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagaca 152
          ||||| | || | | | ||  ||| |  | |  |  | ||| |  | |  | |  |   |
Db 104763 ATGCAACCCAGTAACAAATGATGTTGCAGCAACTACTACAATACACGAGTATACCAAGAA 104704

Qy    153 aatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacg 199
          ||||   |  |||||| || | | |    || ||||| |||||||||
Db 104703 CATCAACAACAACAACTGCAACGGCAACCAACAACAAATTGTATACG 104657


RESULT  12
AAV40300
ID   AAV40300 standard; cDNA; 1751 BP.
XX
AC   AAV40300;
XX
DT   15-OCT-1998  (first entry)
XX
DE   Sugarbeet choline monooxygenase encoding cDNA.
XX
KW   Sugarbeet; choline monooxygenase; CMO; transgenic plant; crop;
KW   stress tolerance; glycine betaine; pathogen; nutrition; ss.
XX
OS   Beta vulgaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             37..1377
FT                   /*tag=  a
FT                   /product= "choline monooxygenase"
XX
PN   WO9830702-A2.
XX
PD   16-JUL-1998.
XX
PF   08-JAN-1998;   98WO-US00164.
XX
PR   08-JAN-1997;   97US-0035147.
XX
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Burnet M,  Hanson AD,  Rathinasabapathi B;
XX
DR   WPI; 1998-399147/34.
DR   P-PSDB; AAW69563.
XX
PT   New polynucleotide molecule encoding choline mono:oxidase - used to,
PT   e.g. transform plants to provide increased resistance to stressful
PT   conditions, increase pest resistance or increase feed value
XX
PS   Claim 45; Fig 4; 36pp; English.
XX
CC   The present sequence encodes sugarbeet choline monooxygenase (CMO). A
CC   process has been developed for increasing the resistance of plants to
CC   stressful conditions, comprising incorporating a nucleotide sequence
CC   encoding CMO into a genome of the plant where the polynucleotide is
CC   expressed and the plant produces glycine betaine. A method has also been
CC   developed of decreasing the production of glycine betaine in a plant,
CC   comprising incorporating a nucleotide sequence encoding CMO in antisense
CC   form into the genome of the plant where the antisense polynucleotide is
CC   expressed and the plant produces no or less glycine betaine. The
CC   expression of CMO in plants can provide plants with resistance to
CC   stressful conditions by osmotic adjustment and osmoprotection. The
CC   blocking of the synthesis of glycine betaine can be used to improve a
CC   plant's resistance to pests or pathogens. Blocking CMO in crop species
CC   used as animal feed may also improve their nutritional value. The
CC   blocking of glycine betaine synthesis can also facilitate industrial
CC   processing, as in sugarbeet, where glycine betaine is an unwanted side
CC   product interfering with sugar crystallisation. The methods can be used
CC   with plants such as tobacco, wheat, barley, corn, sugarcane, sugar beet,
CC   spinach, cotton, sunflower, rice, sorghum, tomato, potato, lettuce,
CC   oilseed rape, or citrus.
XX
SQ   Sequence 1751 BP; 547 A; 344 C; 336 G; 524 T; 0 other;


  Query Match             6.3%;  Score 34.8;  DB 19;  Length 1751;
  Best Local Similarity   48.1%;  Pred. No. 2.4;
  Matches   99;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;
```

```
Qy    274 aatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtcct 333
          || ||   |  || ||| |||| |  || | | |||||||  |  | ||      |  ||
Db    643 aaagccagcaaagcaactgaaacacaaaatttggatcctaaagaacttgggcttgcaccc 702

Qy    334 gtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtt 393
          ||| || | ||| |||| |           ||| || ||    | |||||  |   | |
Db    703 ctaaaagttgcagaatggggcccattcattcttatcagcttggaccgatctctagatgct 762

Qy    394 tcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctg 453
              || ||| |   || |   | | || | | |    |     |  |||  |  |
Db    763 aatgctgatgttggaacagagtggattggtaaatctgcagaagatgttaaggcccatgct 822

Qy    454 tatgacaacaatctgatcaataccaa 479
          | |||    ||||| |     ||| |
Db    823 tttgatcctaatctaaagttcaccca 848
```

RESULT 13
AAV74746/c
ID AAV74746 standard; DNA; 1072 BP.
XX
AC AAV74746;
XX
DT 16-MAR-1999 (first entry)
XX
DE Staphylococcus aureus contig SEQ ID #435.
XX
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX
OS Staphylococcus aureus.
XX
PN EP786519-A2.
XX
PD 30-JUL-1997.
XX
PF 07-JAN-1997; 97EP-0100117.
XX
PR 05-JAN-1996; 96US-0009861.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX
DR WPI; 1997-374922/35.
XX
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
PS Claim 1; Page 1338; 3271pp; English.
XX
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX
SQ Sequence 1072 BP; 376 A; 145 C; 221 G; 326 T; 4 other;

```
  Query Match             6.2%;  Score 34.4;  DB 18;  Length 1072;
  Best Local Similarity   60.9%;  Pred. No. 2.6;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy    329 gtcctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcat 388
          || ||| || | |  ||||| |   ||||||  |||||  ||||||  || |  ||||
Db    275 GTTCTGAAATATAATCAGCAGTGGCTAACGCTGACGTTACAACCTCTTCTGGTTTAGCAT 216

Qy    389 ccgtttcctcttatgctaatactgcaggaaat 420
          |  | | |  ||    ||||  | |||  | |
Db    215 CAATGTGCATTTCACGTAATTTTTCAGTTACT 184
```

RESULT 14
ABA18492/c
ID ABA18492 standard; DNA; 1570 BP.
XX
AC ABA18492;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 10823.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 10823; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1570 BP; 509 A; 230 C; 234 G; 597 T; 0 other;

Query Match 6.2%; Score 34.4; DB 22; Length 1570;
Best Local Similarity 45.5%; Pred. No. 3.1;
Matches 122; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186

```
          |||  || |  |    |  ||||  || |      ||| ||     |  | | |  | |
Db    666 TATTAAACAATCACACATTTGAGGTAATTAGCTGGGAAAAAAAAAAAAAAAATATATATAT 607

Qy    187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
          || | || ||  | || || |  |||| |||| |  ||      |    |||  | |
Db    606 ATATATATACACCTATACATAGACACATCCACAGAATATTCACAAATCTATTGGCCCCAA 547

Qy    247 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 306
          |||      ||| |  ||    || ||||   |  |  |  ||| | ||||||    |
Db    546 CACCAGCAGATAGTCATATTAAATGAAAAATATCACATACCAACAGGAAAAAGTTACACA 487

Qy    307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
          |      |        |   ||| || |||    ||    | |||| |    |||  ||
Db    486 TAGGGAATATACAAAAAATATGGTGCTATAATCCTTAATATAAGGAGTATGTTTTGAATT 427

Qy    367 aacacctcttgcggatctgcatccgttt 394
          ||    |   |   |  ||  ||| |||
Db    426 AATGTTTTGAGGAAAAGTGTTTCCATTT 399


RESULT  15
AAS31317/c
ID   AAS31317 standard; cDNA; 3943 BP.
XX
AC   AAS31317;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human cDNA encoding a novel extracellular matrix protein, Seq ID No 131.
XX
KW   Human; secreted extracellular matrix protein; ss; immunomodulatory;
KW   Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW   antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
KW   human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW   cancers; hyperproliferative disorder; breast neoplasm; melanoma;
KW   Sezary syndrome; Gaucher's disease; neurological diseases;
KW   Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW   cardiac arrest; tachycardia; angina; infection; corneal infections;
KW   wound healing; immunogen; gene therapy; antisense; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155368-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01348.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
```

PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465572/50.
DR P-PSDB; AAU19746.
XX
PT Nucleic acid molecules encoding human secreted extracellular matrix
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
XX
PS Claim 1; SEQ ID No 131; 577pp; English.
XX
CC The invention relates to isolated nucleic acid molecules encoding
CC novel human secreted extracellular matrix proteins (SPs). The
CC polynucleotides and proteins are used to prevent, treat a medical
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. For example, disorders associated with decreased
CC expression of SPs. The SP polynucleotide or a vector expressing them may
CC be administered to treat diseases by gene therapy. Antisense molecules
CC may be administered to down regulate expression of SPs by binding with
CC the cells own genes and preventing their expression. The polynucleotides
CC may also be used as DNA probes in diagnostic assays. The SPs may also be
CC used as antigens to produce antibodies and to identify modulators
CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
CC antagonists may also be used to down regulate expression and activity of
CC SP and as diagnostic agents for detecting the presence of SPs in samples.
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
CC arrest, tachycardia and angina), infections caused by bacteria, viruses
CC and fungi and ocular disorders (e.g. corneal infections). Other uses
CC include wound healing, maintenance of organs before transplantation,

Query Match 6.2%; Score 34.4; DB 22; Length 3943;
Best Local Similarity 53.0%; Pred. No. 4.6;
Matches 71; Conservative 1; Mismatches 62; Indels 0; Gaps 0;

```
Qy      98 tcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagacaaatct 157
           |:  |||||  | | | |       ||      |||||||   |  ||  ||| |||  |
Db    1087 TYCAATTCATTTTTACCGGGAGAAAATACATAGCAAGAGGTAAACAGTGGAGCCAAGACC 1028

Qy     158 taatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaaggcatcaa 217
           |||| || || | |||  ||||  | ||       || | |  ||  | ||  | |  ||
Db    1027 TAATCAATAATGCCTAGCACTGAAATAATGGTAAAGTCTTCTAAATGCCAATAGTACTAA 968

Qy     218 caggaacttgctgt 231
           |||  ||||| | |
Db     967 CAGAGACTTGATAT 954
```

Search completed: June 27, 2002, 20:21:50
Job time: 6408 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 17:35:57 ; Search time 103.12 Seconds
(without alignments)
1322.019 Million cell updates/sec

Title: US-09-727-769A-5
Perfect score: 555
Sequence: 1 ttggcgagtgtaattcctga.....atgtatccagctgtggattt 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		344.2	62.0	555	4	US-09-324-910-5	Sequence 5, Appli
2		344.2	62.0	1080	4	US-09-324-910-11	Sequence 11, Appl
c 3		35.6	6.4	2489	3	US-09-141-047-7	Sequence 7, Appli
4		34.8	6.3	1751	4	US-09-004-393B-3	Sequence 3, Appli
5		34.2	6.2	3666	2	US-08-682-517-13	Sequence 13, Appl
6		34.2	6.2	3666	2	US-08-682-517-14	Sequence 14, Appl
7		34.2	6.2	4197	2	US-08-682-517-7	Sequence 7, Appli
8		34.2	6.2	4197	2	US-08-682-517-8	Sequence 8, Appli
c 9		33.6	6.1	4223	4	US-09-541-782-5	Sequence 5, Appli
c 10		32.8	5.9	5502	3	US-08-836-134-1	Sequence 1, Appli
11		32.4	5.8	7286	3	US-09-331-581-3	Sequence 3, Appli
12		32.4	5.8	7938	3	US-09-331-581-14	Sequence 14, Appl
13		31.8	5.7	1189	4	US-08-936-165A-195	Sequence 195, App
14		31.6	5.7	1810	3	US-08-946-026-19	Sequence 19, Appl
15		31.2	5.6	657	3	US-08-721-979A-74	Sequence 74, Appl
16		31.2	5.6	657	3	US-08-836-501-74	Sequence 74, Appl
17		31.2	5.6	726	3	US-08-836-501-78	Sequence 78, Appl
18		31.2	5.6	1050	3	US-08-836-501-76	Sequence 76, Appl
19		31.2	5.6	1071	3	US-08-836-501-77	Sequence 77, Appl
c 20		31	5.6	1892	2	US-08-933-750C-66	Sequence 66, Appl
c 21		31	5.6	1892	3	US-09-234-613-66	Sequence 66, Appl
c 22		31	5.6	87350	3	US-08-781-891-79	Sequence 79, Appl
c 23		30.4	5.5	7218	1	US-08-232-463-14	Sequence 14, Appl
c 24		30.2	5.4	1007	1	US-07-721-761A-37	Sequence 37, Appl
c 25		30.2	5.4	1007	1	US-07-978-687-37	Sequence 37, Appl
c 26		30.2	5.4	1007	5	PCT-US91-05801-37	Sequence 37, Appl
c 27		30.2	5.4	1573	1	US-07-721-761A-36	Sequence 36, Appl
c 28		30.2	5.4	1573	1	US-07-978-687-36	Sequence 36, Appl
c 29		30.2	5.4	1573	5	PCT-US91-05801-36	Sequence 36, Appl
c 30		29.6	5.3	3126	3	US-08-898-978-1	Sequence 1, Appli
c 31		29.4	5.3	443	1	US-08-592-406-13	Sequence 13, Appl
32		29.2	5.3	5433	3	US-08-929-329-1	Sequence 1, Appli
33		29	5.2	1418	1	US-08-391-615-7	Sequence 7, Appli
34		29	5.2	1823	1	US-08-145-995A-1	Sequence 1, Appli
35		29	5.2	1823	1	US-08-451-747-1	Sequence 1, Appli
36		29	5.2	1823	3	US-09-134-852-1	Sequence 1, Appli
c 37		28.8	5.2	5116	1	US-08-038-682-1	Sequence 1, Appli
c 38		28.8	5.2	5116	1	US-08-302-832-1	Sequence 1, Appli
c 39		28.8	5.2	5116	2	US-08-530-198-1	Sequence 1, Appli
c 40		28.8	5.2	5116	2	US-08-469-880-1	Sequence 1, Appli
c 41		28.8	5.2	5116	2	US-08-728-470-1	Sequence 1, Appli
c 42		28.8	5.2	5116	2	US-08-617-697-1	Sequence 1, Appli
c 43		28.8	5.2	5116	4	US-08-719-641-1	Sequence 1, Appli
44		28.8	5.2	7218	1	US-08-232-463-14	Sequence 14, Appl
c 45		28.4	5.1	1455	3	US-09-009-494-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-324-910-5
; Sequence 5, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Chryseobacterium gleum
US-09-324-910-5

Query Match 62.0%; Score 344.2; DB 4; Length 555;
Best Local Similarity 76.7%; Pred. No. 5.4e-96;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy       7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
           ||||| |||||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db       7 agtgttattcctgatctggcaacgctgaacagtttatttacccagatcaaaaaccaggct 66

Qy      67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
           ||||| || ||||| || || || || || || ||||| ||||||||||| || |||||||||
Db      67 tgcggaacttctacagcatcttctccttgtatcaccttcagatatccggttgacggatgt 126

Qy     127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
           |||||||| || || || |||||||||||| || ||||| ||||||||||||||||||
Db     127 tatgcaagggctcacaaaatgagacaaatcctattgaacgccggctatgactgtgaaaag 186

Qy     187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
           || || ||||| || || || |  || || |||||||| |||||||| | |||  ||
Db     187 cagttcgtatatggtaatctgagagcttctacaggaacatgctgtgtatcatgggtatat 246

Qy     247 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 306
           ||||| ||||| |||||||||| |||||||||| ||| |    |||||||||| ||
Db     247 cacgtagcaattttggtaagcttcaaaaatgcttcaggaattgttgaaaaaagaatcata 306

Qy     307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
           ||||||||| |||| || |||||||||||||||||| | |||||||||   || ||
Db     307 gatccttcattattctccagcggtcctgtaacagattctgcatggagagctgcatgtacc 366
```

```
Qy    367 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
          |||||    |||||||||||| || || || || || || ||||| ||||||||||||||
Db    367 aacacaagctgcggatctgcgtctgtatcttcctacgccaatacagcaggaaatgtttac 426

Qy    427 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 486
          ||||||||||      ||| |  |||||||| |||||     | ||||||| ||||||
Db    427 tacagaagtccgtcaggttcattactgtatgataacaactatgtgaataccaattgtgta 486

Qy    487 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 546
           | |  | ||| |||  ||||| |||||||||||||| || |||||   |||| | |||
Db    487 ttaaacatattctcatccctttcaggatgttctccttccccagcaccaagtgtagcaagc 546

Qy    547 tgtggattt 555
          |||||||||
Db    547 tgtggattt 555


RESULT   2
US-09-324-910-11
; Sequence 11, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
;  APPLICANT: Yamaguchi, Shotaro
;  APPLICANT: Matsuura, Akira
;  TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
;  TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
;  FILE REFERENCE: Q-54362
;  CURRENT APPLICATION NUMBER: US/09/324,910
;  CURRENT FILING DATE: 1999-06-03
;  EARLIER APPLICATION NUMBER: HEI-10-173940
;  EARLIER FILING DATE: 1998-06-04
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Chryseobacterium gleum
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: (61)..(462)
;   FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: (463)..(1017)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (61)..(1017)
US-09-324-910-11


  Query Match           62.0%;  Score 344.2;  DB 4;  Length 1080;
  Best Local Similarity  76.7%;  Pred. No. 7.1e-96;
  Matches  421;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy      7 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 66
          ||||| |||||||||  | || ||  | ||   |||||| |  || || || || ||| ||
Db    469 agtgttattcctgatctggcaacgctgaacagtttatttacccagatcaaaaaccaggct 528

Qy     67 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 126
          ||||| || ||||| || || || || || ||||| |||||||||| || |||||||||
Db    529 tgcggaacttctacagcatcttctccttgtatcaccttcagatatccggttgacggatgt 588

Qy    127 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 186
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    589 tatgcaagggctcacaaaatgagacaaatcctattgaacgccggctatgactgtgaaaag 648

Qy    187 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 246
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    649 cagttcgtatatggtaatctgagagcttctacaggaacatgctgtgtatcatgggtatat 708

Qy    247 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 306
          ||||| ||||| |||||||||| |||||||||| ||| |    |||||||||| ||
Db    709 cacgtagcaattttggtaagcttcaaaaatgcttcaggaattgttgaaaaaagaatcata 768

Qy    307 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 366
          ||||||||| |||| || ||||||||||||||||| | |||||||||   || ||
Db    769 gatccttcattattctccagcggtcctgtaacagattctgcatggagagctgcatgtacc 828

Qy    367 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
          |||||    |||||||||| || || || || || || ||||| ||||||||||||||
Db    829 aacacaagctgcggatctgcgtctgtatcttcctacgccaatacagcaggaaatgtttac 888

Qy    427 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 486
          ||||||||||      ||| |  |||||||| |||||     | ||||||| ||||||
Db    889 tacagaagtccgtcaggttcattactgtatgataacaactatgtgaataccaattgtgta 948

Qy    487 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 546
           | |  | ||| |||  ||||| |||||||||||||| || |||||   |||| | |||
Db    949 ttaaacatattctcatccctttcaggatgttctccttccccagcaccaagtgtagcaagc 1008

Qy    547 tgtggattt 555
          |||||||||
Db   1009 tgtggattt 1017


RESULT   3
US-09-141-047-7/c
; Sequence 7, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
;  APPLICANT: Walker, David H.
;  APPLICANT: Yu, Xue-Jie
;  TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
;  TITLE OF INVENTION: Protein Gene
;  FILE REFERENCE: D6143
;  CURRENT APPLICATION NUMBER: US/09/141,047A
;  CURRENT FILING DATE: 1998-08-27
;  NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 7
;   LENGTH: 2489
;   TYPE: DNA
;   ORGANISM: Ehrlichia canis
;   FEATURE:
;   LOCATION: -340..2149
;   OTHER INFORMATION: Nucleotide sequence of gene encoding 120kDa
;   OTHER INFORMATION: immunoreactive protein
US-09-141-047-7


  Query Match           6.4%;  Score 35.6;  DB 3;  Length 2489;
  Best Local Similarity  45.9%;  Pred. No. 0.21;
  Matches  122;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy    276 tgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtcctgt 335
          | |||||  | ||     | | |   ||  |    |||||||   |  | |   | | |
Db   1843 TTCTTCCTTACTAGTTTCAGATACTTTTTCTCCAACTTCACTTGATGAATGTTCTACACT 1784

Qy    336 aacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtttc 395
          | ||  ||||||| |  | ||  ||| || |   || ||  | | || | | || |
Db   1783 ACCATCTACAGCAGGTTGCAAATCTTCCGCTTTAACTTCAGGAGTATTTTCCTCTTTACT 1724

Qy    396 ctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctgta 455
             ||  | || |  | |   || |     | |   | || |||      | |
Db   1723 AGTTTCAGATACTTTTTCTCCAACTTCACTTGATGAATGTTCTACACTACCATCTACAGC 1664

Qy    456 tgacaacaatctgatcaataccaactgtgtactgactaaattttcactgctttccggatg 515
            |    |||          |   |  |  | | |   |   | || |||  |||| |
Db   1663 AGGTTGCAAATCTTCTGCTTTAACTTCAGGAGTATTTTCCTCTTTACTAGTTTCAGATAC 1604

Qy    516 ttctccttcacctgcaccggatgtat 541
          || | || || || |||  |||| ||
Db   1603 TTTTTCTCCAACTTCACTTGATGAAT 1578


RESULT   4
US-09-004-393B-3
```

```
; Sequence 3, Application US/09004393B
; Patent No. 6310271
; GENERAL INFORMATION:
;  APPLICANT:  Hanson D., Andrew
;  APPLICANT:  Rathinasabapathi, Bala
;  APPLICANT:  Burnet, Michael
;  TITLE OF INVENTION:  Polynucleotides Encoding Choline Monooxygenase and
;  TITLE OF INVENTION:  Plants Transformed Therewith
;  FILE REFERENCE:  UF-162
;  CURRENT APPLICATION NUMBER: US/09/004,393B
;  CURRENT FILING DATE:  1998-01-08
;  PRIOR APPLICATION NUMBER:  60/035,147
;  PRIOR FILING DATE:  1997-01-08
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1751
;   TYPE: DNA
;   ORGANISM: Beta vulgaris
US-09-004-393B-3


  Query Match             6.3%;  Score 34.8;  DB 4;  Length 1751;
  Best Local Similarity   48.1%;  Pred. No. 0.32;
  Matches   99;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy    274 aatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtcct 333
          || ||   |  || ||| |||| |  || | | ||||||  |  | ||      |  ||
Db    643 aaagccagcaaagcaactgaaacacaaaatttggatcctaaagaacttgggcttgcaccc 702

Qy    334 gtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtt 393
          ||| || | ||| |||| |           ||| || ||    | |||||  |   | |
Db    703 ctaaaagttgcagaatggggcccattcattcttatcagcttggaccgatctctagatgct 762

Qy    394 tcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctg 453
              || ||| |   || |   | | || | | |    |     |  |||  |  |
Db    763 aatgctgatgttggaacagagtggattggtaaatctgcagaagatgttaaggcccatgct 822

Qy    454 tatgacaacaatctgatcaataccaa 479
          | |||    ||||| |     ||| |
Db    823 tttgatcctaatctaaagttcaccca 848


RESULT   5
US-08-682-517-13
; Sequence 13, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3666 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-682-517-13


  Query Match             6.2%;  Score 34.2;  DB 2;  Length 3666;
  Best Local Similarity   47.4%;  Pred. No. 0.66;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    272 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 331
          ||  | ||   | ||||||  |   |   || |  ||||| |||   ||| |||   |||
Db   1013 AAGGTACTGATGAAGTAACTGCTTATGCAACTGGTGATCGTTCTAAATTCTCACTTGGTT 1072

Qy    332 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 391
           ||||         | |   | ||    ||  ||| |              |  ||| || |
Db   1073 ATGTATTCTGGGGTGTAGATACAATTCTTTCAGTTGAAGAAGTAACTACAGGTGCTTCAG 1132

Qy    392 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 451
          ||     |  ||| || |   |    || |||  ||| | ||  ||||  | |  | | |
Db   1133 TTAATAATGGTGCAAACAAAACTTACAAAGTTACTTATAAAAACCCTAAAACTGGTAAAC 1192

Qy    452 tgtatgacaacaatctgatcaataccaactgtgta 486
            | |  |||||     | |||      | ||||
Db   1193 CAGAAGCAAACAAAACATTTAATGTTGGTTTTGTA 1227


RESULT   6
US-08-682-517-14
; Sequence 14, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3666 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3666
US-08-682-517-14


  Query Match             6.2%;  Score 34.2;  DB 2;  Length 3666;
  Best Local Similarity   47.4%;  Pred. No. 0.66;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    272 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 331
          ||  | ||   | ||||||  |   |   || |  ||||| |||   ||| |||   |||
Db   1013 AAGGTACTGATGAAGTAACTGCTTATGCAACTGGTGATCGTTCTAAATTCTCACTTGGTT 1072

Qy    332 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 391
           ||||         | |   | ||    ||  ||| |              |  ||| || |
Db   1073 ATGTATTCTGGGGTGTAGATACAATTCTTTCAGTTGAAGAAGTAACTACAGGTGCTTCAG 1132

Qy    392 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 451
          ||     |  ||| || |   |    || |||  ||| | ||  ||||  | |  | | |
Db   1133 TTAATAATGGTGCAAACAAAACTTACAAAGTTACTTATAAAAACCCTAAAACTGGTAAAC 1192

Qy    452 tgtatgacaacaatctgatcaataccaactgtgta 486
            | |  |||||     | |||      | ||||
Db   1193 CAGAAGCAAACAAAACATTTAATGTTGGTTTTGTA 1227


RESULT   7
US-08-682-517-7
; Sequence 7, Application US/08682517
; Patent No. 5874267
```

```
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-682-517-7


  Query Match             6.2%;  Score 34.2;  DB 2;  Length 4197;
  Best Local Similarity   47.4%;  Pred. No. 0.7;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy     272 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 331
           ||  | ||   | ||||||  |   |  || |  ||||| |||   ||| |||   |||
Db    1197 AAGGTACTGATGAAGTAACTGCTTATGCAACTGGTGATCGTTCTAAATTCTCACTTGGTT 1256

Qy     332 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 391
            ||||         | |   | ||    ||  ||| |            |  ||| || |
Db    1257 ATGTATTCTGGGGTGTAGATACAATTCTTTCAGTTGAAGAAGTAACTACAGGTGCTTCAG 1316

Qy     392 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 451
           ||     |  ||| || |   |    || |||  ||| | ||  ||||  | |  | | |
Db    1317 TTAATAATGGTGCAAACAAAACTTACAAAGTTACTTATAAAAACCCTAAAACTGGTAAAC 1376

Qy     452 tgtatgacaacaatctgatcaataccaactgtgta 486
           | |  |||||     | |||       | ||||
Db    1377 CAGAAGCAAACAAAACATTTAATGTTGGTTTTGTA 1411


RESULT   8
US-08-682-517-8
; Sequence 8, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacillus sphaericus
;      INDIVIDUAL ISOLATE:  P-1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  95..3850
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  185..3850
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  95..184
US-08-682-517-8


  Query Match             6.2%;  Score 34.2;  DB 2;  Length 4197;
  Best Local Similarity   47.4%;  Pred. No. 0.7;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy     272 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 331
           ||  | ||   | ||||||  |   |  || |  ||||| |||   ||| |||   |||
Db    1197 AAGGTACTGATGAAGTAACTGCTTATGCAACTGGTGATCGTTCTAAATTCTCACTTGGTT 1256

Qy     332 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 391
            ||||         | |   | ||    ||  ||| |            |  ||| || |
Db    1257 ATGTATTCTGGGGTGTAGATACAATTCTTTCAGTTGAAGAAGTAACTACAGGTGCTTCAG 1316

Qy     392 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 451
           ||     |  ||| || |   |    || |||  ||| | ||  ||||  | |  | | |
Db    1317 TTAATAATGGTGCAAACAAAACTTACAAAGTTACTTATAAAAACCCTAAAACTGGTAAAC 1376

Qy     452 tgtatgacaacaatctgatcaataccaactgtgta 486
           | |  |||||     | |||       | ||||
Db    1377 CAGAAGCAAACAAAACATTTAATGTTGGTTTTGTA 1411


RESULT   9
US-09-541-782-5/c
; Sequence 5, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
;  APPLICANT:  Nislow, Corey
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  TITLE OF INVENTION:  Antifungal Assay
;  FILE REFERENCE:  1015
;  CURRENT APPLICATION NUMBER: US/09/541,782
;  CURRENT FILING DATE:  2000-04-03
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 4223
;   TYPE: DNA
;   ORGANISM: S.pombe
US-09-541-782-5


  Query Match             6.1%;  Score 33.6;  DB 4;  Length 4223;
  Best Local Similarity   52.1%;  Pred. No. 1.1;
  Matches   75;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     336 aacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtttc 395
           | ||   | | | |   | |   | | | ||| | |||  || |  ||| |||  || |
Db    3073 AGCATCAATATCGTTACGCACACCATACATTAGCTCCTTCTGTGATTCTTCATGAGTATG 3014

Qy     396 ctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctgta 455
           || | |     ||| ||     |||   |     ||  |||    | |||||| | |||
Db    3013 CTGTAAATTATATAGTGATCTTAATTCCTGAACAAGTTGTCGGTATGATTCTTTCATGTC 2954

Qy     456 tgacaacaatctgatcaataccaa 479
           |  |  ||||   |||||| || |
Db    2953 TTTCTGCAATTCTATCAATTCCGA 2930


RESULT  10
US-08-836-134-1/c
; Sequence 1, Application US/08836134A
; Patent No. 6020127
```

```
; GENERAL INFORMATION:
;  APPLICANT: Mackenzie, Alex E.
;  APPLICANT: Korneluk, Robert G.
;  APPLICANT: Mahadevan, Mani S.
;  APPLICANT: McLean, Michael
;  APPLICANT: Roy, Natalie
;  APPLICANT: Ikeda, Joh-e
;  TITLE OF INVENTION: Neuronal Apoptosis Inhibitor Protein, Gene Sequence and
; Patent No. 6020127
;  TITLE OF INVENTION: Mutations Causative of Spinal Muscular Atrophy
;  FILE REFERENCE: 3477-112, 033477/139914
;  CURRENT APPLICATION NUMBER: US/08/836,134A
;  CURRENT FILING DATE: 1997-06-20
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 5502
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-836-134-1


  Query Match             5.9%;  Score 32.8;  DB 3;  Length 5502;
  Best Local Similarity   58.0%;  Pred. No. 2.1;
  Matches   58;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy    238 tggagctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaa 297
          |||| | |  |  |  | |||| |   |  | |  |||| |||| ||| | || ||||||
Db   3805 TGGATCAATAATTTCTCCATATGGTGGAAGTTTGGAAATTCTTCAGGAATGACTGAAAAA 3746

Qy    298 agaattattgatccttcactattttcaagcggtcctgtaa 337
          | | ||||    || ||   ||   ||  | || || | |
Db   3745 ACATTTATATTGCCCTCCAGATCCACAGACAGTTCTTTCA 3706


RESULT  11
US-09-331-581-3
; Sequence 3, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
;  APPLICANT: TOHDA, Hideki
;  APPLICANT: HAMA, Yuko
;  APPLICANT: KUMAGAI, Hiromicho
;  TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
;  TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
;  TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
;  FILE REFERENCE: 0059-1142-0PCT
;  CURRENT APPLICATION NUMBER: US/09/331,581
;  CURRENT FILING DATE: 1999-06-30
;  EARLIER APPLICATION NUMBER: PCT/JP98/04929
;  EARLIER FILING DATE: 1998-10-30
;  EARLIER APPLICATION NUMBER: JP 9-314608
;  EARLIER FILING DATE: 1997-10-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7286
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3


  Query Match             5.8%;  Score 32.4;  DB 3;  Length 7286;
  Best Local Similarity   48.4%;  Pred. No. 3.1;
  Matches   90;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy    117 agacggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatga 176
          ||| |  | |   ||   |    | | || |  |||| |  ||  || ||     |  |
Db   4574 agatgtctatgtagctgaaaataaaatgacgtcacaagacaaaaaaaaaaaaaaaaaaa 4633

Qy    177 ctgtgaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggc 236
          ||||| ||     | |   |||   |||   |  || | | || | ||| |  |
Db   4634 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaagtaccttctgaggcgg 4693

Qy    237 gtggagctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaa 296
          || | | |  | | ||    || |||| || |   |||  |  |||  |||   ||
Db   4694 aaagaaccagccggatccagacatgataagatacattgatgagtttggacaaaccacaac 4753

Qy    297 aagaat 302
          ||||||
Db   4754 tagaat 4759


RESULT  12
US-09-331-581-14
; Sequence 14, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
;  APPLICANT: TOHDA, Hideki
;  APPLICANT: HAMA, Yuko
;  APPLICANT: KUMAGAI, Hiromicho
;  TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
;  TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
;  TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
;  FILE REFERENCE: 0059-1142-0PCT
;  CURRENT APPLICATION NUMBER: US/09/331,581
;  CURRENT FILING DATE: 1999-06-30
;  EARLIER APPLICATION NUMBER: PCT/JP98/04929
;  EARLIER FILING DATE: 1998-10-30
;  EARLIER APPLICATION NUMBER: JP 9-314608
;  EARLIER FILING DATE: 1997-10-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 7938
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-14


  Query Match             5.8%;  Score 32.4;  DB 3;  Length 7938;
  Best Local Similarity   48.4%;  Pred. No. 3.2;
  Matches   90;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy    117 agacggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatga 176
          ||| |  | |   ||   |    | | || |  |||| |  ||  || ||     |  |
Db   5226 agatgtctatgtagctgaaaataaaatgacgtcacaagacaaaaaaaaaaaaaaaaaaa 5285

Qy    177 ctgtgaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggc 236
          ||||| ||     | |   |||   |||   |  || | | || | ||| |  |
Db   5286 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaagtaccttctgaggcgg 5345

Qy    237 gtggagctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaa 296
          || | | |  | | ||    || |||| || |   |||  |  |||  |||   ||
Db   5346 aaagaaccagccggatccagacatgataagatacattgatgagtttggacaaaccacaac 5405

Qy    297 aagaat 302
          ||||||
Db   5406 tagaat 5411


RESULT  13
US-08-936-165A-195
; Sequence 195, Application US/08936165A
; Patent No. 6348582
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Burnham, Martin
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Lonetto, Michael
;    APPLICANT:  Nicholas, Richard
```

```
;    APPLICANT:  Pratt, Julie
;    APPLICANT:  Reichard, Richard
;    APPLICANT:  Rosenberg, Martin
;    APPLICANT:  Ward, Judith
;    TITLE OF INVENTION:  No. 6348582el Prokaryotic Polynucleotides,
;    TITLE OF INVENTION:  Polypeptides and Their Uses
;    NUMBER OF SEQUENCES:  534
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,165A
;      FILING DATE:  24-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,032
;      FILING DATE:  24-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  195:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1189 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-936-165A-195


  Query Match             5.7%;  Score 31.8;  DB 4;  Length 1189;
  Best Local Similarity   56.1%;  Pred. No. 2.3;
  Matches   60;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     159 aatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaaggcatcaac 218
           |||| ||| || || | ||||| |  | |   | ||| |  ||| || ||   || |||
Db     726 AATGGACAGCGCCTTTAACTGTTGATGAAACAGTTTATGCACAAAGCTTAACAGATAAAC 785

Qy     219 aggaacttgctgtgtggcgtggagctaccacgttgcaatattggtaa 265
             |     | | |||    |||| ||   ||  | | | | |||| |
Db     786 CAGTTAAAGGTATGTTAACTGGACCTGTAACAATTCTAAACTGGTCA 832


RESULT  14
US-08-946-026-19
; Sequence 19, Application US/08946026
; Patent No. 6034218
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Mitcham, Jennifer L.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;    TITLE OF INVENTION:  AND IMMUNODIAGNOSIS OF PROSTATE CANCER
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/946,026
;      FILING DATE:  07-OCT-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.424C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1810 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-946-026-19


  Query Match             5.7%;  Score 31.6;  DB 3;  Length 1810;
  Best Local Similarity   54.2%;  Pred. No. 3.1;
  Matches   64;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      97 atcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagacaaatc 156
           | || |   ||| ||| |||       ||   ||  |||||    |   | || | ||| |
Db    1419 AACAGAGCAAGAAATCATGTCAGTTTATAAAATCAAAGAGAGACTTCTAAGACAAAAAGC 1478

Qy     157 ttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaaggcat 214
           | || ||  | |   |||||| ||||| | |  ||  | |  |  | | | ||| ||
Db    1479 ATCATTAATGAGGCAGATGACTCTGAAAGAAACATTTAAAAAAGCCATCCAGAGGAAT 1536


RESULT  15
US-08-721-979A-74
; Sequence 74, Application US/08721979A
; Patent No. 6113911
;  GENERAL INFORMATION:
;    APPLICANT:  Binz, Hans
;    APPLICANT:  N'Guyen, Ngoc Thien
;    APPLICANT:  Baussant, Thierry
;    APPLICANT:  Trudel, Michel
;    TITLE OF INVENTION:  PEPTIDE FRAGMENT OF RESPIRATORY
;    TITLE OF INVENTION:  SYNCYTIAL VIRUS PROTEIN G, IMMUNOGENIC AGENT,PHARMACEUTICAL
;    TITLE OF INVENTION:  COMPOSITION CONTAINING IT AND PREPARATION PROCESS
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gordon W. Hueschen
;      STREET:  715 The "H" Bldg., 310 East Michigan
;      STREET:  Avenue
;      CITY:  Kalamazoo
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  49007
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      SOFTWARE:
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/721,979A
;      FILING DATE:  October 4, 1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  FR 94 04009
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hueschen, Gordon W.
;      REGISTRATION NUMBER:  16,157
;      REFERENCE/DOCKET NUMBER:  PF57PCTUS/dln
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  616-382-0030
;      TELEFAX:  616-382-2030
;  INFORMATION FOR SEQ ID NO:  74:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  657 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..657
;      OTHER INFORMATION:  /product= "BB"
US-08-721-979A-74


  Query Match             5.6%;  Score 31.2;  DB 3;  Length 657;
  Best Local Similarity   49.4%;  Pred. No. 2.7;
  Matches   81;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy    116 tagacggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatg 175
          | |||  || |   | ||| | | || | | ||  | ||||   |    |||      ||
Db    446 TTGACAAATATGGAGTAAGTGACTATTACAAGAACCTAATCAACAATGCCAAAACTGTTG 505

Qy    176 actgtgaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtgg 235
          |  ||| ||||  | |  || | | ||   ||   |||| | |    ||| .|  |
Db    506 AAGGTGTAAAAGCACTGATAGATGAAATTTTAGCTGCATTACCTAAGACTGACACTTACA 565

Qy    236 cgtggagctaccacgttgcaatattggtaagctataaaaatgct 279
            |  | |    | | |  || ||||  | || | | || | ||
Db    566 AATTAATCCTTAATGGTAAAACATTGAAAGGCGAAACAACTACT 609


Search completed: June 27, 2002, 20:15:14
Job time: 9557 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 16:25:12 ; Search time 2534.16 Seconds
(without alignments)
2955.934 Million cell updates/sec

Title: US-09-727-769A-5
Perfect score: 555
Sequence: 1 ttggcgagtgtaattcctga.....atgtatccagctgtggattt 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	40.8	7.4	606	12	FR0009895	AL001177 F.rubripe
	2	40.4	7.3	233	9	AA767965	AA767965 nv96a12.s
	3	39	7.0	532	10	BI893381	BI893381 sai65d11.
c	4	39	7.0	570	9	AW937387	AW937387 QV3-DT004
c	5	39	7.0	570	9	AW937408	AW937408 QV3-DT004
c	6	37.6	6.8	858	10	BF697486	BF697486 602129379
	7	37.4	6.7	531	10	BM092866	BM092866 saj02b03.
c	8	37.2	6.7	728	12	AG164728	AG164728 Pan trogl
	9	36.8	6.6	550	12	AQ982897	AQ982897 RPCI-23-2
c	10	36.6	6.6	555	10	BG683527	BG683527 602651387
	11	36.4	6.6	491	12	AQ143297	AQ143297 HS_3067_B
	12	36.2	6.5	668	10	BI183607	BI183607 UNL-P-FN-
c	13	35.6	6.4	642	12	AZ525610	AZ525610 246PbA04
c	14	35.6	6.4	686	12	AQ390063	AQ390063 RPCI11-15
c	15	35.6	6.4	1101	12	CNS0100X	AL098379 Drosophil
	16	35.4	6.4	625	9	AI722559	AI722559 fd19g10.x
c	17	35.4	6.4	894	12	AZ531159	AZ531159 ENTBN36TR
c	18	35.4	6.4	938	12	CNS006TJ	AL065906 Drosophil
	19	35.2	6.3	635	12	AQ982969	AQ982969 RPCI-23-2
	20	35	6.3	488	10	BI366489	BI366489 RE52036.5
c	21	35	6.3	539	12	AQ708947	AQ708947 HS_2121_A
	22	35	6.3	666	10	BI240734	BI240734 RE37648.5
	23	35	6.3	687	9	AV682345	AV682345 AV682345
c	24	35	6.3	1001	12	CNS012LO	AL101718 Drosophil
	25	34.8	6.3	367	10	BE581125	BE581125 kq42f01.y
	26	34.8	6.3	434	10	BE579597	BE579597 kq31c03.y
	27	34.8	6.3	465	10	BE579625	BE579625 kq31e10.y
c	28	34.8	6.3	590	12	AZ285039	AZ285039 RPCI-23-1
	29	34.8	6.3	600	12	AZ005132	AZ005132 RPCI-23-3
	30	34.8	6.3	685	12	AQ995355	AQ995355 RPCI-23-2
	31	34.8	6.3	759	9	AI541345	AI541345 pec1.2-7.
c	32	34.6	6.2	533	12	AQ413419	AQ413419 RPCI-11-1
	33	34.4	6.2	271	12	BH188461	BH188461 036_E_02-
	34	34.4	6.2	271	12	CNS07SH2	AL625400 T3 end of
c	35	34.4	6.2	567	9	AW398915	AW398915 EST309415
	36	34.2	6.2	501	12	AZ905458	AZ905458 RPCI-24-1
c	37	34.2	6.2	978	12	CNS01PQH	AL155145 Anopheles
c	38	34	6.1	156	9	AW164768	AW164768 se77e05.y
c	39	34	6.1	351	9	AW216307	AW216307 687047G07
	40	34	6.1	448	12	AZ802912	AZ802912 2M0063D04
	41	34	6.1	477	12	BH634370	BH634370 SALK_0452
	42	34	6.1	564	10	BF296339	BF296339 035PbF05
c	43	34	6.1	571	10	BM023631	BM023631 ie82h05.y
	44	34	6.1	664	12	AQ356929	AQ356929 CITBI-E1-
	45	34	6.1	846	10	BM398346	BM398346 5009-0-44

ALIGNMENTS

RESULT 1
FR0009895/c
LOCUS FR0009895 606 bp DNA linear GSS 18-SEP-1997
DEFINITION F.rubripes GSS sequence, clone 032C09aD10, genomic survey sequence.
ACCESSION AL001177
VERSION AL001177.1 GI:2439029
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 606)
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y., Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic sequence.
FEATURES Location/Qualifiers
source 1. .606
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 032C09"
/clone="032C09aD10"
BASE COUNT 165 a 150 c 110 g 160 t 21 others
ORIGIN

Query Match 7.4%; Score 40.8; DB 12; Length 606;
Best Local Similarity 50.3%; Pred. No. 0.63;
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 183 aaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggag 242

```
          | || ||||   ||  ||  ||  ||  |  || |    |  ||   |   |   | |
Db    553 AGAATAATTGTAATGTGGTNACAGAAGTGATTCCATCCCACATTTTNGAACGTNTTTGGT 494

Qy    243 ctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaat 302
          | | ||||| | | | |||||| || |||  ||   |  ||   | ||   ||||
Db    493 AAGGTATGCTGCAAAAGTCGAAAGCTAAAATAATAATTATCGTATATGCGCAAGTNGAAT 434

Qy    303 tattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttg 362
          |  |  ||| |   |||| | ||      |||||  |  || |   | || || |||  |
Db    433 TTATACTCCATTCTTATTNTGAAAATTCTCTGTAGGAAGTAGATGGTCGATAAGCGCGCG 374

Qy    363 cgtta 367
          | | |
Db    373 CTTGA 369


RESULT   2
AA767965
LOCUS       AA767965               233 bp    mRNA    linear   EST 07-FEB-1998
DEFINITION  nv96a12.s1 NCI_CGAP_Pr16 Homo sapiens cDNA clone IMAGE:1237630,
            mRNA sequence.
ACCESSION   AA767965
VERSION     AA767965.1  GI:2818980
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 332   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1237630"
                     /clone_lib="NCI_CGAP_Pr16"
                     /sex="male"
                     /tissue_type="tumor"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pAMP10; mRNA made from
                     invasive prostate tumor cells, cDNA made by  oligo-dT
                     priming.  Non-directionally cloned.  Size-selected on
                     agarose  gel, average insert size 600 bp.  Library made by
                     D. Krizman, NIH."
BASE COUNT       84 a     41 c     40 g     68 t
ORIGIN


  Query Match             7.3%;  Score 40.4;  DB 9;  Length 233;
  Best Local Similarity   49.5%;  Pred. No. 0.59;
  Matches  104;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy     18 tgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttgcggtacctc 77
          || | |   || | || | ||| ||  || | |   | ||      ||||     ||||
Db     21 TGGTTTGTTTAAACTACAATCTATACACAGTTCACATATGACATTTACTTGTTAGGCCTC 80

Qy     78 tacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgttatgcaagagc 137
          |       |   || ||||    |  |     | ||| | |||  ||||     | ||||
Db     81 TTTTTGTCCAAAACGATGCGGTTCCATACAGCAGCCTCTGGACACATGTAGCTGATGAGC 140

Qy    138 ccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttgtata 197
            |||| ||  | | | ||  ||| ||     ||| | | | | |||  |||  || ||
Db    141 AGATAAAATATGGCTAGTCCAAATTAAGGTATGCTGTAAGTATAAAATGCAAAGTGGATT 200

Qy    198 cggaaacctaaaggcatcaacaggaacttg 227
            ||| |   |  ||  || |  || | |
Db    201 TCCAAAACAGTATTCAAAAAAAAAAAGTCG 230


RESULT   3
BI893381
LOCUS       BI893381               532 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  sai65d11.y1 Gm-c1068 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1068-3621 5', mRNA sequence.
ACCESSION   BI893381
VERSION     BI893381.1  GI:16105641
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1068-3621"
                     /clone_lib="Gm-c1068"
                     /tissue_type="Leaf, drought stressed, 1 month old plants,
                     greenhouse grown"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from drought stressed leaf tissue of the cultivar Williams
                     82. The month old greenhouse grown plants were deprived of
                     water for 3 days prior to harvesting the stressed leaf
                     tissue. Complementary DNA was synthesized from mRNA using
                     a primer consisting of a poly(dT) sequence with a XhoI
                     restriction site. EcoRI adapters were ligated to the
                     blunt-ended cDNA fragments followed by XhoI digestion. The
                     cDNA fragments were directionally cloned into the
                     EcoRI-XhoI restriction site of the pBluescript vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (GibcoBRL). This library was constructed in the
                     laboratory of Dr. Randy Shoemaker."
BASE COUNT      160 a    115 c    112 g    145 t
ORIGIN
```

```
  Query Match             7.0%;  Score 39;  DB 10;  Length 532;
  Best Local Similarity   68.4%;  Pred. No. 2;
  Matches   54;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      34 aattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctcacca 93
           |||  ||| ||| |   || || | || | |    |||| | ||||||| ||| ||||||
Db       5 AATCTTTTGTTCCAAACAAAAACGTATGAATTGGTCGGTTCTTCTACGGAGTCGTCACCA 64

Qy      94 tgcatcacattcagatatc 112
           || ||||  ||||| || |
Db      65 TGAATCATCTTCAGTTACC 83


RESULT   4
AW937387/c
LOCUS       AW937387                 570 bp    mRNA    linear   EST 30-MAY-2000
DEFINITION  QV3-DT0043-180100-061-c07 DT0043 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW937387
VERSION     AW937387.1  GI:8112810
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV3-DT0043-180
            100-061-c07&t3=2000-01-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 569.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="DT0043"
                     /dev_stage="Adult"
                     /note="Organ: denis_drash; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      203 a     83 c     94 g    190 t
ORIGIN


  Query Match             7.0%;  Score 39;  DB 9;  Length 570;
  Best Local Similarity   52.8%;  Pred. No. 2;
  Matches   84;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     182 aaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgga 241
           ||||  |  || |||| | ||| ||     | ||   |        || | |||  |  |
Db     480 AAAAGGATGTTTTATAGGAAAATCTGGTTACTTCTTGATATTAGAACTTTATGGATTAAA 421

Qy     242 gctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaa 301
           ||||||  |||   ||   | || |||| ||  |  |    |||  || ||  |||
Db     420 GCTACCTTGTTATTATGGGGTTAGGCTAGAAGTAGCCCAAAGGAAAAAAGGGTAAATTTG 361

Qy     302 ttattgatccttcactattttcaagcggtcctgtaacag 340
           ||  ||   ||||| |  || |||  ||| |  || ||
Db     360 CTAAAGAAATTTCACAAAATTTAAGTTGTCTTTAAAAAG 322


RESULT   5
AW937408/c
LOCUS       AW937408                 570 bp    mRNA    linear   EST 30-MAY-2000
DEFINITION  QV3-DT0043-190100-061-c07 DT0043 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW937408
VERSION     AW937408.1  GI:8112831
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV3-DT0043-190
            100-061-c07&t3=2000-01-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 569.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="DT0043"
                     /dev_stage="Adult"
                     /note="Organ: denis_drash; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      203 a     83 c     94 g    190 t
ORIGIN


  Query Match             7.0%;  Score 39;  DB 9;  Length 570;
  Best Local Similarity   52.8%;  Pred. No. 2;
  Matches   84;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     182 aaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgga 241
           ||||  |  || |||| | ||| ||     | ||   |        || | |||  |  |
Db     480 AAAAGGATGTTTTATAGGAAAATCTGGTTACTTCTTGATATTAGAACTTTATGGATTAAA 421

Qy     242 gctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaa 301
           ||||||  |||   ||   | || |||| ||  |  |    |||  || ||  |||
```

```
Db    420 GCTACCTTGTTATTATGGGGTTAGGCTAGAAGTAGCCCAAAGGAAAAAAGGGTAAATTTG 361

Qy    302 ttattgatccttcactattttcaagcggtcctgtaacag 340
          ||  ||   ||||| |  || |||  ||| |  || ||
Db    360 CTAAAGAAATTTCACAAAATTTAAGTTGTCTTTAAAAAG 322


RESULT   6
BF697486/c
LOCUS       BF697486                858 bp    mRNA    linear   EST 22-DEC-2000
DEFINITION  602129379F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4286394 5',
            mRNA sequence.
ACCESSION   BF697486
VERSION     BF697486.1  GI:11982894
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1122  row: m  column: 19
            High quality sequence stop: 537.
FEATURES             Location/Qualifiers
     source          1. .858
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4286394"
                     /clone_lib="NIH_MGC_56"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      247 a    168 c    151 g    292 t
ORIGIN


  Query Match             6.8%;  Score 37.6;  DB 10;  Length 858;
  Best Local Similarity   47.5%;  Pred. No. 5.9;
  Matches  112;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy    191 ttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctaccacg 250
          |||| |  | ||    | ||  ||  || | || ||  ||| |  |     |   | |
Db    314 TTGTGTTGGTAATTTAACAGTAATATACTGAAAATTAGTGTATAACTCCAGGAACCAATA 255

Qy    251 ttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattattgatc 310
          ||  | | ||||||  |  |||| |||   | ||   |   | ||    || | |   |
Db    254 ATGGTACAGTGGTAAAATGGAAAAGTGCCCACTGAAATAAATATAATTTCATCACTTCCC 195

Qy    311 cttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgttaaca 370
          |    |||      | ||  |    | || ||||   |||||     |      | |
Db    194 TTCAGAAATTCCACCACTGTGTTCATTCGTATTCAGCTGAAAGAAAAATAAGGACATAAA 135

Qy    371 cctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 426
          ||   | |  || || || | ||| || | | |   ||  |  | | |||| || |
Db    134 CCCTATACTTATTTGGATGCTTTTGCTGTAAGGTATATTTTAAATGTAATGATTTT 79


RESULT   7
BM092866
LOCUS       BM092866                531 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  saj02b03.y1 Gm-c1065 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1065-9269 5', mRNA sequence.
ACCESSION   BM092866
VERSION     BM092866.1  GI:17021832
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1065-9269"
                     /clone_lib="Gm-c1065"
                     /tissue_type="germinating shoots"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     germinating shoots of the cultivar Williams. The seeds
                     were allowed to germinate for 24  hours prior to being
                     cold stressed for 2 days at 4C. Complementary DNA  was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT)  sequence with a XhoI restriction site. EcoRI
                     adapters were ligated to  the blunt-ended cDNA fragments
                     followed by XhoI digestion. The cDNA  fragments were
                     directionally cloned into the EcoRI-XhoI restriction  site
                     of the pBluescript vector. The ligated cDNA fragments were
                     transformed into DH10B host cells (GibcoBRL). This library
                     was  constructed in the laboratory of Dr. Randy
                     Shoemaker."
BASE COUNT      153 a    127 c     87 g    164 t
ORIGIN


  Query Match             6.7%;  Score 37.4;  DB 10;  Length 531;
  Best Local Similarity   67.1%;  Pred. No. 5.7;
  Matches   53;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy     34 aattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctcacca 93
          |||  ||| ||| |   || || | || | |    ||||   |||||| ||| ||||||
Db     91 AATCTTTTGTTCCAAACAAAAACGTATGAATTGGTCGGTTTTTCTACGGAGTCGTCACCA 150
```

```
Qy      94 tgcatcacattcagatatc 112
           || ||||  ||||| || |
Db     151 TGAATCATCTTCAGTTACC 169


RESULT   8
AG164728/c
LOCUS       AG164728                728 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-031P16.TJ, genomic survey
            sequence.
ACCESSION   AG164728
VERSION     AG164728.1  GI:16694406
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-031P16.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 728)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .728
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-031P16.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      164 a    188 c    169 g    205 t      2 others
ORIGIN


  Query Match             6.7%;  Score 37.2;  DB 12;  Length 728;
  Best Local Similarity   50.0%;  Pred. No. 7.2;
  Matches   93;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy     123 atgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtga 182
           |||| || || ||||   |   | || |||   || |||    | | |   || | ||
Db     448 ATGTCATTCAGGAGCATGTCTAAGGACACACGGCTAAATAGTGAGCTGTAGTGGCAGTAT 389

Qy     183 aaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggag 242
             ||      || | |  ||    | | |||    |      |   || | | |||||
Db     388 TCAAGCTCAGGTCTGGGTGAATTATATGCCATTCTAAATGTGTGAATGAGCTGTGTGGAC 329

Qy     243 ctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaat 302
           |||  |||      |  |||| ||| | |  ||||  ||||   | || |  || ||  |
Db     328 CTAATACGGGTGCCTCCTGGTGAGCCAAAGTAATGGGTCCGATCTCACAGCCAACAGTGT 269

Qy     303 tattga 308
           |  |||
Db     268 TGGTGA 263
```

```
RESULT   9
AQ982897
LOCUS       AQ982897                550 bp    DNA     linear   GSS 30-JAN-2000
DEFINITION  RPCI-23-262D23.TJ RPCI-23 Mus musculus genomic clone RPCI-23-262D23
            , DNA sequence.
ACCESSION   AQ982897
VERSION     AQ982897.1  GI:6816102
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-262D23.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 262  row: D  column: 23
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-262D23"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      181 a    122 c    109 g    137 t      1 others
ORIGIN


  Query Match             6.6%;  Score 36.8;  DB 12;  Length 550;
  Best Local Similarity   53.9%;  Pred. No. 8.6;
  Matches   97;  Conservative    0;  Mismatches   82;  Indels    1;  Gaps    1;

Qy      40 ttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctcaccatgcatc 99
           |||  ||| |||| ||| || ||  |    |  |     || ||  || || |||  | |
Db     201 TTAAACAAACAAACAAACAAACAACAGTAATGATCACAGCAACCACAGCATCGAAC 260

Qy     100 acattcagatatcctgtagacggatgttatgcaagagcccataagatgagacaaatctta 159
           |||   | | ||  ||| |  |  || | |  || ||   | |  ||||||||||
Db     261 ACACAAACAAATGGTGTGGCAGAGTGGTGTAAAAAAGGAGAGATTATGAGACAAAATAGG 320

Qy     160 atgaacaacggctatga-ctgtgaaaaacaatttgtatacggaaacctaaaggcatcaac 218
           | | |   |||||   || | ||| |   ||  |||   |||    | | |||||| |
Db     321 CAGTAAATGTGCTATCTTCTTTTAAAGAACCTTAATATTGAGAATGGCATAAGCATCATC 380
```

```
RESULT 10
BG683527/c
LOCUS       BG683527                555 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  602651387T1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4762196 3',
            mRNA sequence.
ACCESSION   BG683527
VERSION     BG683527.1  GI:13914924
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1615  row: n  column: 21
            High quality sequence start: 37
            High quality sequence stop: 555.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4762196"
                     /clone_lib="NIH_MGC_47"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      160 a    116 c     97 g    182 t
ORIGIN


  Query Match             6.6%;  Score 36.6;  DB 10;  Length 555;
  Best Local Similarity   50.3%;  Pred. No. 9.8;
  Matches   90;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy     339 agatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtttcctc 398
           | ||| |||| ||||||||||     | |  |||   | | || | | || |||    |
Db     286 AAATAAAGCAACGAGAAACGCTATTAATGATGCCTTCAGTGTATATTCCTCTGTTGTTCC 227

Qy     399 ttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctgtatga 458
           | ||||| | | |  ||| |    | | |   | | ||    ||| | | |    | ||
Db     226 TAATGCTCAAAATCAAGGGACCTCTGAATGGTGTACTCTGGCTAAATGTAAGACATCTGG 167

Qy     459 caacaatctgatcaataccaactgtgtactgactaaattttcactgctttccggatgtt 517
           || || |   | || |   |    |  | | |    | | || |   ||| |  | ||
Db     166 CATCACTTGCAGCACTGTAACACCTTCAGTCTCAGTCTGTGCAATTACTTCTGTTTCTT 108


RESULT 11
AQ143297
LOCUS       AQ143297                491 bp    DNA     linear   GSS 24-SEP-1998
DEFINITION  HS_3067_B1_A11_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3067 Col=21 Row=B, DNA sequence.
ACCESSION   AQ143297
VERSION     AQ143297.1  GI:3533950
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3067  row: B  column: 21
            Class: BAC ends
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3067 Col=21 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      169 a     93 c     90 g    131 t      8 others
ORIGIN


  Query Match             6.6%;  Score 36.4;  DB 12;  Length 491;
  Best Local Similarity   50.3%;  Pred. No. 11;
  Matches   88;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy     152 aaatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaagg 211
           ||  |  ||||   | | |||   |  |  |   | |||    |||| | |  |     |
Db     316 AACACCAAATGTGTATCTGCTGGAAGGGACATGCAGAATACACATACAGTACTCAGCGTG 375

Qy     212 catcaacaggaacttgctgtgtggcgtggagctaccacgttgcaatattggtaagctata 271
             | ||  ||    |   |   ||| || |||  | ||  | ||  ||| |     |  || |
Db     376 AAACACCAAATGCACATTTGCTGGAGTTGAGAAATCAGGCTGGAATTTGTANGACGTAAA 435

Qy     272 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaag 326
           | ||||  ||   ||   |   |  ||||    |||||||| ||   | |||||
Db     436 ACAATGGCTCTACAGACCCATGACCAAGATGACTTGATCCTTAACACGTCTCAAG 490


RESULT 12
BI183607
LOCUS       BI183607                668 bp    mRNA    linear   EST 10-JUL-2001
DEFINITION  UNL-P-FN-ca-e-10-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-ca-e-10-0-UNL 3', mRNA sequence.
ACCESSION   BI183607
VERSION     BI183607.1  GI:14658016
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
```

```
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-ca-e-10-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected during 7 days of the follicular phase
                     of the pig estrous cycle. This library was derived from
                     the library UNL-P-F2. The tag is a string of 5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track. The library was constructed as described
                     by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      173 a    163 c    150 g    176 t      6 others
ORIGIN


  Query Match             6.5%;  Score 36.2;  DB 10;  Length 668;
  Best Local Similarity   57.5%;  Pred. No. 14;
  Matches   65;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    178 tgtgaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcg 237
          |||| |  | ||| ||  |  |  ||   | | | ||||  | |||| || ||| ||||
Db    248 TGTGCACCAGAATCTGCCTGGGCTAATGGATATGGATCAGGATGAACGTGTTGTTTGGCC 307

Qy    238 tggagctaccacgttgcaatattggtaagctataaaaatgcttccggagtaac 290
          | ||| | |    || | |||||   |     || ||| ||||| | |  | |
Db    308 TAGAGATTCACTCTTCCGATATTTTGAGTTGCTAGAAAAGCTTCAGTATAACC 360


RESULT  13
AZ525610/c
LOCUS       AZ525610                642 bp    DNA     linear   GSS 07-MAY-2001
DEFINITION  246PbA04 Pb MBN #21 Plasmodium berghei genomic 3', DNA sequence.
ACCESSION   AZ525610
VERSION     AZ525610.1  GI:13966638
KEYWORDS    GSS.
SOURCE      Plasmodium berghei.
  ORGANISM  Plasmodium berghei
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   Carlton,J.M.-R. and Dame,J.B.
  TITLE     The Plasmodium vivax and P. berghei gene sequence tag projects
  JOURNAL   Parasitol. Today 16 (10), 409 (2000)
COMMENT     Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Plasmodium berghei"
                     /strain="ANKA clone 15cy1 (clone of the ANKA 8417 clone)"
                     /db_xref="taxon:5821"
                     /clone_lib="Pb MBN #21"
                     /dev_stage="asexual blood forms"
                     /lab_host="Mus musculus"
                     /note="Vector: pBluescript SK(+) vector DNA, phagemid
                     excised from lambda ZAP; Site_1: EcoRV; Site_2: EcoRV;
                     Genomic DNA was prepared from asynchronous blood stage
                     forms of the cloned ANKA isolate of P. berghei grown in
                     laboratory Swiss white mice. The DNA was purified from
                     contaminating host DNA by Hoechst Dye 33258-CsCl
                     ultracentrifugation and precipitated. Purified DNA was
                     digested with mung bean nuclease in the presence of 36-38%
                     formamide at 50 C, as described (Vernick, K.D., Imberski,
                     R.B., and McCutchan, T.F. 1988. Nucleic Acids Research
                     16:6883-6896). The ends of the digestion fragments were
                     polished using T4 DNA polymerase, and the fragments size
                     selected in the range 500-2000 bp. These were ligated into
                     the EcoRV-cleaved and dephosphorylated pBluescript SK(+)
                     vector. Recombinant plasmids were used to transform E.
                     coli XL10-Gold host cells. "
BASE COUNT      194 a     98 c     79 g    270 t      1 others
ORIGIN


  Query Match             6.4%;  Score 35.6;  DB 12;  Length 642;
  Best Local Similarity   58.5%;  Pred. No. 20;
  Matches   62;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    107 gatatcctgtagacggatgttatgcaagagcccataagatgagacaaatcttaatgaaca 166
          | |||  | |  | | ||  ||| | | |    | ||   ||||  ||    ||  || |
Db    272 GGTATATTTTTTATGAATTATATACCATATTGAAAAAATGGAGAACAAAAACAAGTAATA 213

Qy    167 acggctatgactgtgaaaaacaatttgtatacggaaacctaaaggc 212
          |    |||| |||||||||||||| |   |||   ||||  ||||||
Db    212 AACTGTATGTCTGTGAAAAACAAATGTGATAAATAAACACAAAGGC 167


RESULT  14
AQ390063/c
LOCUS       AQ390063                686 bp    DNA     linear   GSS 21-MAY-1999
DEFINITION  RPCI11-155D7.TV RPCI-11 Homo sapiens genomic clone RPCI-11-155D7,
            DNA sequence.
ACCESSION   AQ390063
VERSION     AQ390063.1  GI:4361086
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 686)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-155D7.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
```

```
FEATURES             Location/Qualifiers
     source          1. .686
                     /organism="Homo sapiens"
                     /db_xref="GDB:7559214"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-155D7"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      217 a    123 c     85 g    261 t
ORIGIN
```

Query Match 6.4%; Score 35.6; DB 12; Length 686;
Best Local Similarity 47.0%; Pred. No. 20;
Matches 143; Conservative 0; Mismatches 159; Indels 2; Gaps 1;

```
Qy       6 gagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtc 65
Db     506 GGGTTTAAATCTGCTTTTTACTCTTTGTTATTTGTATTTATTTGAAAATATCTATTGGTC 447

Qy      66 ttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatg 125
Db     446 TTGTTTTTCACCTATAAAATTATAATACTTAATAACAAGCCTTTAAAATGTTGTAGTCCA 387

Qy     126 ttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaa 185
Db     386 TTATTATAAATGTAAAAGTACATTACATATTTTAAGTCCAAGCAAATATTAGTGTTATAC 327

Qy     186 a--caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagc 243
Db     326 ATCTAGATAGGGAAGAGAGAAAGCAAGGTATACGAAAAGACTTCCATTTAGGATTTGATG 267

Qy     244 taccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaatt 303
Db     266 TAACTAACTGCCATTTAGTTAATTTTCTTAGAAGTGTAAGGAATGACTTAAAGAAAAACT 207

Qy     304 attg 307
Db     206 TGTG 203
```

```
RESULT 15
CNS0100X/c
LOCUS       CNS0100X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098379
VERSION     AL098379.1  GI:5609990
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03G04"
                     /note="end : SP6"
BASE COUNT      195 a    108 c    131 g    161 t    506 others
ORIGIN
```

Query Match 6.4%; Score 35.6; DB 12; Length 1101;
Best Local Similarity 17.8%; Pred. No. 24;
Matches 89; Conservative 191; Mismatches 218; Indels 1; Gaps 1;

```
Qy      30 attaaattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctc 89
Db    1067 HTHTHHHRHTHWHHHTWYTMTATTHHKWMWTYNCRYKCHYNMYNYMTCHMTCCMYCHCWC 1008

Qy      90 accatgcatcacattcagatatcctgtagacggatgttatgcaagagcccataagatgag 149
Db    1007 CMCWCWHAWMWMWHMWHWAWAWVTHTWTWMWWMCCTCMCTMCMYCYYYCTCTYCYCYYNY 948

Qy     150 acaaatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaa 209
Db     947 CMWAAAHHACAKAAMWAMSHWMYMWCWCHCHTMMMHCTMHACTCAHCHCTMMAYMMYMYM 888

Qy     210 ggcatcaacaggaacttgctgtgtggcgtggagcta-ccacgttgcaatattggtaagct 268
Db     887 HMYYYCYMYCYTYMYYCTCYCHCYCCCCYCYMVHMATMTMCAMTATAMTWATWTACAHHT 828

Qy     269 ataaaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcg 328
Db     827 WWWTAMDHMMMTMMMMCTYHMWCWMYAHHHYHYCTMWHMHTHHWHCHMHCHYMTMHHHHY 768

Qy     329 gtcctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcat 388
Db     767 HTMYTAAAMMAAAAAAAAAAAAMAAAAAAAAMMAAAMATWTATMTTATTTATHTTTTYATMT 708

Qy     389 ccgtttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattctt 448
Db     707 YWTWTYHCMTMSHCHCCACMHMYHCHCYMCHYMMHAWMHMWCWMAHACAMAMHAMMMMAM 648

Qy     449 acctgtatgacaacaatctgatcaataccaactgtgtactgactaaattttcactgcttt 508
Db     647 AAMMWCCMMCWCWCYMYCTMAHTHMWHWHMWHWWHTHAHHAAHTMAAACWCHAMATHAHT 588

Qy     509 ccggatgttctccttcacc 527
Db     587 MCHTHMHMTNTTYHTYTCH 569
```

Search completed: June 27, 2002, 19:22:05
Job time: 10613 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 20:15:32 ; Search time 84.27 Seconds
(without alignments)
243.843 Million cell updates/sec

Title: US-09-727-769A-6
Perfect score: 991
Sequence: 1 LASVIPDVATLNSLFNQIKN.....FSLLSGCSPSPAPDVSSCGF 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	991	100.0	185	22	AAB84386	Amino acid sequenc
2	991	100.0	320	22	AAB84387	Amino acid sequenc
3	874	88.2	185	21	AAY44582	Chryseobacterium g
4	874	88.2	319	21	AAY44583	Chryseobacterium g
5	108	10.9	20	22	AAB84385	Internal peptide o
6	96	9.7	20	22	AAB84384	N-terminal peptide
7	91	9.2	594	21	AAG52451	Arabidopsis thalia
8	91	9.2	594	21	AAG52479	Arabidopsis thalia
9	85.5	8.6	477	22	AAB30815	Amino acid sequenc
10	85	8.6	264	21	AAG25396	Arabidopsis thalia
11	85	8.6	292	21	AAG25395	Arabidopsis thalia
12	85	8.6	294	21	AAG25394	Arabidopsis thalia
13	84	8.5	3726	22	ABB63947	Drosophila melanog
14	82	8.3	927	20	AAY35054	Chlamydia pneumoni
15	82	8.3	928	20	AAW88423	Chlamydia pneumoni
16	82	8.3	928	21	AAY90239	Chlamydia antigen
17	81	8.2	270	21	AAG21040	Arabidopsis thalia
18	81	8.2	304	21	AAG21039	Arabidopsis thalia
19	81	8.2	346	22	ABB71526	Drosophila melanog
20	81	8.2	486	21	AAG21038	Arabidopsis thalia
21	81	8.2	1137	14	AAR42081	Impatiens Necrotic
22	79	8.0	1290	22	ABB62818	Drosophila melanog
23	76.5	7.7	433	13	AAR26274	Alkali-protease Ya
24	76.5	7.7	433	19	AAW61495	Modified Bacillus
25	76.5	7.7	433	20	AAW95698	Bacillus sp. Lion
26	76.5	7.7	433	21	AAY69207	Amino acid sequenc
27	76.5	7.7	433	21	AAY44619	Bacillus Lion Y en
28	76.5	7.7	499	13	AAR27574	The ABF-B from A.
29	76.5	7.7	636	20	AAW89548	Bacillus sp. alkal
30	76.5	7.7	726	21	AAB36463	Human plakophilin
31	76.5	7.7	849	17	AAW06725	FLO1 protein, invo
32	76	7.7	1237	18	AAW55640	H. pylori ORF 04ep
33	76	7.7	1237	20	AAY17187	H. pylori outer me
34	75.5	7.6	244	16	AAR70731	Human protein S de
35	75.5	7.6	244	16	AAR70732	Human protein S de
36	75.5	7.6	244	16	AAR70733	Human protein S de
37	75.5	7.6	390	16	AAR70730	Human protein S de
38	75.5	7.6	468	21	AAB25882	Synthetic 55kD i-a
39	75.5	7.6	565	16	AAR70729	Human protein S de
40	75.5	7.6	675	16	AAR70728	Human protein S.
41	75.5	7.6	1219	22	AAE03801	Human DNA helicase
42	74.5	7.5	763	22	ABB62815	Drosophila melanog
43	74.5	7.5	1047	22	ABB59191	Drosophila melanog
44	74.5	7.5	1247	22	ABB61174	Drosophila melanog
45	74	7.5	211	21	AAG25261	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAB84386
ID AAB84386 standard; Protein; 185 AA.
XX
AC AAB84386;
XX
DT 22-AUG-2001 (first entry)
XX
DE Amino acid sequence of a protein-deamidating enzyme.
XX
KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW bakery; confectionery.
XX
OS Cryseobacterium sp.
XX
PN EP1106696-A1.
XX
PD 13-JUN-2001.
XX
PF 04-DEC-2000; 2000EP-0310768.
XX
PR 03-DEC-1999; 99JP-0345044.
XX
PA (AMAN-) AMANO ENZYME INC.
XX
PI Yamaguchi S;
XX
DR WPI; 2001-376907/40.
DR N-PSDB; AAF90280.
XX
PT New enzyme for use in e.g. bakery has an ability to deamidate amido
PT groups in a protein -
XX
PS Claim 4; Page 22; 43pp; English.

```
XX
CC   The present sequence represents a protein-deamidating enzyme from
CC   Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
CC   groups in a protein by directly acting upon the amido groups without
CC   cutting peptide bonds and without cross-linking the protein. The enzyme
CC   thus reduces the mineral sensitivity of the protein and increases the
CC   soluble mineral content in the protein-mineral solution, improving the
CC   absorption of minerals in the human body. The enzyme reduces or removes
CC   the toxicity and allergenic property of the protein in a food
CC   e.g. allergy. The enzyme is useful for the improvement of dough in the
CC   field of bakery and confectionery e.g. for the production of crackers,
CC   biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC   protein products, in various food articles e.g. meat or fish products
CC   and noodles; and for improving functionality of plant or animal protein.
XX
SQ   Sequence   185 AA;


  Query Match             100.0%;  Score 991;  DB 22;  Length 185;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-94;
  Matches  185;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 lasvipdvatlnslfnqiknqscgtstasspcitfrypvdgcyarahkmrqilmnngydc 60

Qy     61 EKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ekqfvygnlkastgtccvawsyhvailvsyknasgvtekriidpslfssgpvtdtawrna 120

Qy    121 CVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cvntscgsasvssyantagnvyyrspsnsylydnnlintncvltkfsllsgcspspapdv 180

Qy    181 SSCGF 185
          |||||
Db    181 sscgf 185


RESULT   2
AAB84387
ID   AAB84387 standard; Protein; 320 AA.
XX
AC   AAB84387;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Amino acid sequence of a protein-deamidating enzyme.
XX
KW   Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW   bakery; confectionery.
XX
OS   Cryseobacterium sp.
XX
FH   Key             Location/Qualifiers
FT   Region          1..135
FT                   /note= "prepro region"
XX
PN   EP1106696-A1.
XX
PD   13-JUN-2001.
XX
PF   04-DEC-2000; 2000EP-0310768.
XX
PR   03-DEC-1999;   99JP-0345044.
XX
PA   (AMAN-) AMANO ENZYME INC.
XX
PI   Yamaguchi S;
XX
DR   WPI; 2001-376907/40.
DR   N-PSDB; AAF90281.
XX
PT   New enzyme for use in e.g. bakery has an ability to deamidate amido
PT   groups in a protein  -
XX
PS   Example 11; Page 24; 43pp; English.
XX
CC   The present sequence represents a protein-deamidating enzyme from
CC   Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
CC   groups in a protein by directly acting upon the amido groups without
CC   cutting peptide bonds and without cross-linking the protein. The enzyme
CC   thus reduces the mineral sensitivity of the protein and increases the
CC   soluble mineral content in the protein-mineral solution, improving the
CC   absorption of minerals in the human body. The enzyme reduces or removes
CC   the toxicity and allergenic property of the protein in a food
CC   e.g. allergy. The enzyme is useful for the improvement of dough in the
CC   field of bakery and confectionery e.g. for the production of crackers,
CC   biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC   protein products, in various food articles e.g. meat or fish products
CC   and noodles; and for improving functionality of plant or animal protein.
XX
SQ   Sequence   320 AA;


  Query Match             100.0%;  Score 991;  DB 22;  Length 320;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-94;
  Matches  185;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 lasvipdvatlnslfnqiknqscgtstasspcitfrypvdgcyarahkmrqilmnngydc 195

Qy     61 EKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 ekqfvygnlkastgtccvawsyhvailvsyknasgvtekriidpslfssgpvtdtawrna 255

Qy    121 CVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 cvntscgsasvssyantagnvyyrspsnsylydnnlintncvltkfsllsgcspspapdv 315

Qy    181 SSCGF 185
          |||||
Db    316 sscgf 320


RESULT   3
AAY44582
ID   AAY44582 standard; Protein; 185 AA.
XX
AC   AAY44582;
XX
DT   04-APR-2000  (first entry)
XX
DE   Chryseobacterium gleum protein-deamidating enzyme.
XX
KW   Protein deamidating enzyme; soil bacterium; deamidation activity;
KW   amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW   protein engineering; surface hydrophobicity; toxicity; allergic;
KW   mineral sensitivity; calcium; absorption; mineral enhancing agent.
XX
OS   Chryseobacterium gleum 'JCM 2410'.
XX
PN   EP976829-A2.
XX
PD   02-FEB-2000.
XX
PF   04-JUN-1999;   99EP-0304367.
XX
PR   04-JUN-1998;   98JP-0173940.
XX
PA   (AMAN ) AMANO PHARM KK.
XX
PI   Yamaguchi S,  Matsuura A;
XX
DR   WPI; 2000-118552/11.
```

```
DR   N-PSDB; AAZ49494.
XX
PT   New enzyme for modifying and improving the function of proteins and/or
PT   peptides has deamidating activity without causing cross linking -
XX
PS   Claim 5; Page 24; 57pp; English.
XX
CC   The present amino acid sequence is the protein-deamidating enzyme,
CC   isolated from a new strain of soil bacterium, Chryseobacterium gleum
CC   JCM 2410. This enzyme exerts the deamidation activity by directly acting
CC   upon side chain amido groups in the protein in bonded state and
CC   releasing side chain carboxyl groups and ammonia. It can deaminate high
CC   molecular weight proteins, without cross linking and cleavage of peptide
CC   bonds, to improve protein function. This sequence is used for protein
CC   engineering, to cause an increase in surface hydrophobicity and improve
CC   the function of a plant or animal protein. It can also be used to remove
CC   or reduce toxicity or allergic property of proteins in food, decrease
CC   mineral sensitivity of protein, to allow greater absorption into the body
CC   and to solubilise calcium for use in drinks and mineral enhancing agents.
XX
SQ   Sequence   185 AA;


  Query Match             88.2%;  Score 874;  DB 21;  Length 185;
  Best Local Similarity   87.4%;  Pred. No. 2.9e-82;
  Matches  160;  Conservative   10;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       3 SVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEK 62
           |||||:|||||||  |||||:|||||||||||||||||||||||||||||:| ||||||
Db       3 svipdlatlnslftqiknqacgtstasspcitfrypvdgcyarahkmrqillnagydcek 62

Qy      63 QFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACV 122
           |||||||:||||||||:| ||||||||:|||||: |||||||||||||||||:||| ||
Db      63 qfvygnlrastgtccvswvyhvailvsfknasgivekriidpslfssgpvtdsawraact 122

Qy     123 NTSCGSASVSSYANTAGNVYYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSS 182
           |||||||||||||||||||||||| | ||||| :||||||  || ||||||||||  |:|
Db     123 ntscgsasvssyantagnvyyyrspsgslllydnnyvntncvlnifsslsgcspspapsvas 182

Qy     183 CGF 185
           |||
Db     183 cgf 185


RESULT   4
AAY44583
ID   AAY44583 standard; Protein; 319 AA.
XX
AC   AAY44583;
XX
DT   04-APR-2000  (first entry)
XX
DE   Chryseobacterium gleum protein-deamidating prepro-enzyme.
XX
KW   Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;
KW   amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW   protein engineering; surface hydrophobicity; toxicity; allergic;
KW   mineral sensitivity; calcium; absorption; mineral enhancing agent.
XX
OS   Chryseobacterium gleum 'JCM 2410'.
XX
FH   Key             Location/Qualifiers
FT   Region          1..134
FT                   /label= Prepro_region
FT                   /note= "Comprises the Pre and Pro region"
FT   Peptide         1..21
FT                   /label= Signal_peptide
FT                   /note= "Corresponds to the Pre region"
FT   Peptide         22..134
FT                   /note= "Corresponds to the Pro region"
FT   Protein         135..319
FT                   /label= Mature_protein_deamidating_enzyme
FT                   /note= "Deamidates amido groups in a protein"
XX
PN   EP976829-A2.
XX
PD   02-FEB-2000.
XX
PF   04-JUN-1999;   99EP-0304367.
XX
PR   04-JUN-1998;   98JP-0173940.
XX
PA   (AMAN ) AMANO PHARM KK.
XX
PI   Yamaguchi S,  Matsuura A;
XX
DR   WPI; 2000-118552/11.
DR   N-PSDB; AAZ49495.
XX
PT   New enzyme for modifying and improving the function of proteins and/or
PT   peptides has deamidating activity without causing cross linking -
XX
PS   Example 26; Page 26; 57pp; English.
XX
CC   The present amino acid sequence is the protein-deamidating
CC   prepro-enzyme, isolated from a new strain of soil bacterium,
CC   Chryseobacterium gleum JCM 2410. This enzyme exerts the deamidation
CC   activity by directly acting upon side chain amido groups in the protein
CC   in bonded state and releasing side chain carboxyl groups and ammonia. It
CC   can deaminate high molecular weight proteins, without cross linking and
CC   cleavage of peptide bonds, to improve protein function. This sequence is
CC   used for protein engineering, to cause an increase in surface
CC   hydrophobicity and improve the function of a plant or animal protein. It
CC   can also be used to remove or reduce toxicity or allergic property of
CC   proteins in food, decrease mineral sensitivity of protein, to allow
CC   greater absorption into the body and to solubilise calcium for use in
CC   drinks and mineral enhancing agents.
XX
SQ   Sequence   319 AA;


  Query Match             88.2%;  Score 874;  DB 21;  Length 319;
  Best Local Similarity   87.4%;  Pred. No. 6e-82;
  Matches  160;  Conservative   10;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       3 SVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEK 62
           |||||:|||||||  |||||:|||||||||||||||||||||||||||||:| ||||||
Db     137 svipdlatlnslftqiknqacgtstasspcitfrypvdgcyarahkmrqillnagydcek 196

Qy      63 QFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACV 122
           |||||||:||||||||:| ||||||||:|||||: |||||||||||||||||:||| ||
Db     197 qfvygnlrastgtccvswvyhvailvsfknasgivekriidpslfssgpvtdsawraact 256

Qy     123 NTSCGSASVSSYANTAGNVYYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSS 182
           |||||||||||||||||||||||| | ||||| :||||||  || ||||||||||  |:|
Db     257 ntscgsasvssyantagnvyyyrspsgslllydnnyvntncvlnifsslsgcspspapsvas 316

Qy     183 CGF 185
           |||
Db     317 cgf 319


RESULT   5
AAB84385
ID   AAB84385 standard; peptide; 20 AA.
XX
AC   AAB84385;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Internal peptide of protein-deamidating enzyme.
XX
KW   Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW   bakery; confectionery.
XX
OS   Cryseobacterium sp.
```

XX
PN EP1106696-A1.
XX
PD 13-JUN-2001.
XX
PF 04-DEC-2000; 2000EP-0310768.
XX
PR 03-DEC-1999; 99JP-0345044.
XX
PA (AMAN-) AMANO ENZYME INC.
XX
PI Yamaguchi S;
XX
DR WPI; 2001-376907/40.
XX
PT New enzyme for use in e.g. bakery has an ability to deamidate amido
PT groups in a protein -
XX
PS Example 11; Page 20; 43pp; English.
XX
CC The present sequence represents a peptide fragment of a
CC protein-deamidating enzyme from Cryseobacterium sp. number 9670. The
CC enzyme is able to deamidate amido groups in a protein by directly
CC acting upon the amido groups without cutting peptide bonds and without
CC cross-linking the protein. The enzyme thus reduces the mineral
CC sensitivity of the protein and increases the soluble mineral content
CC in the protein-mineral solution, improving the absorption of minerals
CC in the human body. The enzyme reduces or removes the toxicity and
CC allergenic property of the protein in a food e.g. allergy. The enzyme
CC is useful for the improvement of dough in the field of bakery and
CC confectionery e.g. for the production of crackers, biscuits, cookies,
CC pizza pies or crusts of pie; in producing soybean protein products,
CC in various food articles e.g. meat or fish products and noodles; and
CC for improving functionality of plant or animal protein.
XX
SQ Sequence 20 AA;

Query Match 10.9%; Score 108; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 SPSNSYLYDNNLINTNCVLT 164
||||||||||||||||||||
Db 1 spsnsylydnnlintncvlt 20

RESULT 6
AAB84384
ID AAB84384 standard; peptide; 20 AA.
XX
AC AAB84384;
XX
DT 22-AUG-2001 (first entry)
XX
DE N-terminal peptide of protein-deamidating enzyme.
XX
KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW bakery; confectionery.
XX
OS Cryseobacterium sp.
XX
PN EP1106696-A1.
XX
PD 13-JUN-2001.
XX
PF 04-DEC-2000; 2000EP-0310768.
XX
PR 03-DEC-1999; 99JP-0345044.
XX
PA (AMAN-) AMANO ENZYME INC.
XX
PI Yamaguchi S;
XX
DR WPI; 2001-376907/40.
XX
PT New enzyme for use in e.g. bakery has an ability to deamidate amido
PT groups in a protein -
XX
PS Example 11; Page 20; 43pp; English.
XX
CC The present sequence represents a peptide fragment of a
CC protein-deamidating enzyme from Cryseobacterium sp. number 9670. The
CC enzyme is able to deamidate amido groups in a protein by directly
CC acting upon the amido groups without cutting peptide bonds and without
CC cross-linking the protein. The enzyme thus reduces the mineral
CC sensitivity of the protein and increases the soluble mineral content
CC in the protein-mineral solution, improving the absorption of minerals
CC in the human body. The enzyme reduces or removes the toxicity and
CC allergenic property of the protein in a food e.g. allergy. The enzyme
CC is useful for the improvement of dough in the field of bakery and
CC confectionery e.g. for the production of crackers, biscuits, cookies,
CC pizza pies or crusts of pie; in producing soybean protein products,
CC in various food articles e.g. meat or fish products and noodles; and
CC for improving functionality of plant or animal protein.
XX
SQ Sequence 20 AA;

Query Match 9.7%; Score 96; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LASVIPDVATLNSLFNQIKN 20
||||||||||||||||||||
Db 1 lasvipdvatlnslfnqikn 20

RESULT 7
AAG52451
ID AAG52451 standard; Protein; 594 AA.
XX
AC AAG52451;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66675.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             9.2%;  Score 91;  DB 21;  Length 594;
  Best Local Similarity   24.6%;  Pred. No. 1.1;
  Matches   31;  Conservative   14;  Mismatches   45;  Indels   36;  Gaps    5;

Qy      25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMNNGY----DCEKQFVYGNLKAST-- 73
           |:   :| :|| |   |||   |      : ::: :| :     |:|| |:  |  |
Db     442 tnvddnpsvtfnyfkhpvdlqrcveairlvskvvtsnrflnytqcdkqnvhkmlslsvka 501

Qy      74 --------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                               |    | ||   ||       :   ||   |:|| | |
Db     502 ninlrpkqlndtksmaqfckdtvvtiwhyhggclvgkvvspnrkvlgvdrlrvidgstfd 561

Qy     109 SGPVTD 114
             | |:
Db     562 espgtn 567


RESULT   8
AAG52479
ID   AAG52479 standard; Protein; 594 AA.
XX
AC   AAG52479;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 66714.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
```

PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.2%; Score 91; DB 21; Length 594;
Best Local Similarity 24.6%; Pred. No. 1.1;
Matches 31; Conservative 14; Mismatches 45; Indels 36; Gaps 5;

```
Qy      25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMNNGY----DCEKQFVYGNLKAST-- 73
           |:   :| :|| |   |||   |      : ::: :| :     |:|| |:  |  |
Db     442 tnvddnpsvtfnyfkhpvdlqrcveairlvskvvtsnrflnytqcdkqnvhkmlslsvka 501

Qy      74 --------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                               |    | ||   ||       :   ||   |:|| | |
Db     502 ninlrpkqlndtksmaqfckdtvvtiwhyhggclvgkvvspnrkvlgvdrlrvidgstfd 561

Qy     109 SGPVTD 114
             | |:
Db     562 espgtn 567
```

RESULT 9
AAB30815
ID AAB30815 standard; Protein; 477 AA.
XX
AC AAB30815;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of a prion-like amyloidogenic protein.
XX
KW SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;
KW aggregation; fibril; phenotypic alteration; gene therapy;
KW disease resistance; plant pigmentation; prion disease.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200075324-A2.
XX
PD 14-DEC-2000.
XX
PF 09-JUN-2000; 2000WO-US15876.
XX
PR 09-JUN-1999; 99US-0138833.
XX
PA (ARCH-) ARCH DEV CORP.
XX
PI Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;
XX
DR WPI; 2001-061723/07.
XX
PT New nucleic acid encoding chimeric proteins with self-assembly
PT properties, useful e.g. for diagnosis and treatment of prion diseases,
PT also related aggregates, fibrils and polymers -
XX
PS Claim 11; Page 162-163; 188pp; English.
XX

```
CC   AAB30803-17 represent prion-like amyloidogenic proteins which have
CC   been identified from a yeast protein database. The specification
CC   describes chimeric polypeptides, which comprise at least one SCHAG
CC   (self-coalesces into higher-order aggregates) amino acid sequence fused
CC   in frame with a polypeptide of interest (which is other than a marker
CC   protein, a glutathione-S-transferase or a staphylococcal nuclear
CC   protein). The specification also describes chimeric polypeptides that
CC   comprises an amyloidogenic domain that causes aggregation into fibrils.
CC   The chimeric polypeptides are used to prepare polymers with multiple
CC   reactivities, e.g. derivatised with enzymes, or specific binding
CC   partners, and useful e.g. for performing multi-step chemical reactions.
CC   They can be used create an inducible, or stable phenotypic alteration
CC   in a cell, e.g. for gene therapy, protein production, imparting disease
CC   resistance to plants, altering plant pigmentation and for diagnosis
CC   and treatment of prion diseases.
XX
SQ   Sequence   477 AA;


  Query Match             8.6%;  Score 85.5;  DB 22;  Length 477;
  Best Local Similarity   26.9%;  Pred. No. 3.1;
  Matches   36;  Conservative   16;  Mismatches   43;  Indels   39;  Gaps    6;

Qy      55 NNGYDCEKQFVYGNLKAS--------TGTCCVAWSYH-----------VAILVS-----Y 90
           ||||       ||:  ::        |||  |  | |            | : |     |
Db     310 nngystssthyyghtssasaaagatgtgtanvvssmhannnsassatstayvysmnvnvy 369

Qy      91 KNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVS-SYANTAGNVYY---RSP 146
            |:|    ||              |:  ||   ||  | ::| :|||:  :|||    :
Db     370 ynssasaykr-----------anttsntnasgatstnsgtmsnayansytsvyygyamas 418

Qy     147 SNSYLYDNNLINTN 160
           :||  : : :  ||
Db     419 ansmyhhhtvyatn 432


RESULT  10
AAG25396
ID   AAG25396 standard; Protein; 264 AA.
XX
AC   AAG25396;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 29443.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
```

PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.6%; Score 85; DB 21; Length 264;
Best Local Similarity 24.6%; Pred. No. 1.6;
Matches 31; Conservative 13; Mismatches 46; Indels 36; Gaps 5;

```
Qy      25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMN----NGYDCEKQFVYGNLKAST-- 73
           |:   :| :|| |   |||   |      : ::: :    |   |:|| |:  |  |
Db     112 tnvddnpsvtfnyfkhpvdlqrcveairlvskvvtskrflnytqcdkqnvhkmlslsvka 171

Qy      74 --------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                               |    | ||   ||       :   ||   |:|| | |
Db     172 ninlrpkqlndtksmaqfckdtvvtiwhyhggclvgkvvspnrkvlgvdrlrvidgstfd 231

Qy     109 SGPVTD 114
             | |:
Db     232 espgtn 237
```

RESULT 11
AAG25395
ID AAG25395 standard; Protein; 292 AA.
XX
AC AAG25395;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 29442.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

```
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             8.6%;  Score 85;  DB 21;  Length 292;
  Best Local Similarity   24.6%;  Pred. No. 1.8;
  Matches   31;  Conservative   13;  Mismatches   46;  Indels   36;  Gaps    5;

Qy      25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMN----NGYDCEKQFVYGNLKAST-- 73
           |:   :| :|| |   |||   |      : ::: :    |   |:|| |:  |  |
Db     140 tnvddnpsvtfnyfkhpvdlqrcveairlvskvvtskrflnytqcdkqnvhkmlslsvka 199

Qy      74 --------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                               |    | ||   ||        :   ||   |:|| | |
Db     200 ninlrpkqlndtksmaqfckdtvvtiwhyhggclvgkvvspnrkvlgvdrlrvidgstfd 259

Qy     109 SGPVTD 114
           | |:
Db     260 espgtn 265


RESULT  12
AAG25394
ID   AAG25394 standard; Protein; 294 AA.
XX
AC   AAG25394;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 29441.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
```

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.6%; Score 85; DB 21; Length 294;
Best Local Similarity 24.6%; Pred. No. 1.8;
Matches 31; Conservative 13; Mismatches 46; Indels 36; Gaps 5;

```
Qy      25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMN----NGYDCEKQFVYGNLKAST-- 73
           |:   :| :|| |   |||   |      : ::: :    |   |:|| |:  |  |
Db     142 tnvddnpsvtfnyfkhpvdlqrcveairlvskvvtskrflnytqcdkqnvhkmlslsvka 201

Qy      74 --------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                               |    | ||   ||       :   ||   |:|| | |
Db     202 ninlrpkqlndtksmaqfckdtvvtiwhyhggclvgkvvspnrkvlgvdrlrvidgstfd 261

Qy     109 SGPVTD 114
             | |:
Db     262 espgtn 267
```

RESULT 13
ABB63947
ID ABB63947 standard; Protein; 3726 AA.
XX
AC ABB63947;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 18633.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL08050.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

```
XX
PS   Disclosure; SEQ ID NO 18633; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   3726 AA;


  Query Match             8.5%;  Score 84;  DB 22;  Length 3726;
  Best Local Similarity   25.3%;  Pred. No. 70;
  Matches   46;  Conservative   21;  Mismatches   73;  Indels   42;  Gaps    7;

Qy      22 SCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEK------QFVYGNLKAST-- 73
           | | |: |||        ||               ::|  | |      :   | |  :|
Db     178 ssgrssgsspdgnsgassdga-----------ssgiscgkstaksteassgklakttga 225

Qy      74 GTCCVAWSYHVAILVSYKNASGVTEKRIID-------PSLFSSGPVTDTAWRNACVNTSC 126
           |||  |         | | :|| | :            :|| : | | |    | |:
Db     226 gtcssa--------ksskassgttseattsglsgaclkalfvatpatstglacalvspg- 276

Qy     127 GSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGC------SPSPAPDV 180
           ||:   ::  :|  :  |  ||    : ||     :|   | |         :|||:| :
Db     277 gssqggtfpisaallrarknsnkkfknlnlargevmlpstsklkqlnspvvdnpspsppi 336

Qy     181 SS 182
           :|
Db     337 as 338


RESULT  14
AAY35054
ID   AAY35054 standard; Protein; 927 AA.
XX
AC   AAY35054;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Chlamydia pneumoniae surface exposed polypeptide.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Page 942-944; Disclosure; 1912pp; English.
XX
CC   AAY34584-Y35879 represent the proteins encoded by all the open reading
CC   frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC   C. pneumoniae causes respiratory disease such as pneumonia and
CC   bronchitis and is thought to be a contributing factor in heart
CC   disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC   nodosum or pharyngitis. The polypeptides encoded by the open reading
CC   frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC   immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC   nucleotides sequences can also be used as immunogenic compositions,
CC   especially where the vector directs the expression of a neutralising
CC   epitope of C. pneumoniae.
XX
SQ   Sequence   927 AA;


  Query Match             8.3%;  Score 82;  DB 20;  Length 927;
  Best Local Similarity   21.8%;  Pred. No. 17;
  Matches   49;  Conservative   35;  Mismatches   69;  Indels   72;  Gaps   13;

Qy      10 TLNSLFNQIKNQSCGTSTA-SSPCITFRYPVDGCYARAHKMRQILMN--NGYDCE----- 61
           |: :|   :   : |: || :: |  |:        : |  : :| |   | :|
Db      57 tiynltgdvsitnagsptaltasc--fkettgnlsfqghgyqfllqnidaganctftnta 114

Qy      62 -------KQFVY------------GNLKASTGTCCVAWSYHVAILVSYKNASG------ 95
                  | |              | :| ||| | :  :|      :: | :|
Db     115 ankllsfsgfsylsliqttnattgtgaik-stgacsiqsnyscyfgqnfsndnggalqgs 173

Qy      96 ----------------VTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVS--SYANT 137
                           |:|     :|:|:| :|        :| :  ||| |  : ||
Db     174 sislslnpnltfaknkatqk---ggalystggit--------inntlnsasfsentaann 222

Qy     138 AGNVYYRSPSNSYLYDNNLIN-TNCVLTKFSLLSG---CSPSPAP 178
           | :|  : ::|::  |  |:  |  :|  |   |   || : ||
Db     223 ggaiy--teassfissnkaisfinnsvtatsatggaiycsstsap 265


RESULT  15
AAW88423
ID   AAW88423 standard; Protein; 928 AA.
XX
AC   AAW88423;
XX
DT   26-APR-1999  (first entry)
XX
DE   Chlamydia pneumoniae surface exposed protein Omp10.
XX
KW   Omp10; outer membrane protein 10; surface exposed protein;
KW   antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9858953-A2.
XX
PD   30-DEC-1998.
XX
PF   19-JUN-1998;   98WO-DK00266.
XX
PR   23-JUN-1997;   97DK-0000744.
XX
PA   (BIRK/) BIRKELUND S.
PA   (CHRI/) CHRISTIANSEN G.
XX
PI   Birkelund S,  Christiansen G,  Knudsen K,  Madsen A;
PI   Mygind P;
XX
DR   WPI; 1999-105610/09.
DR   N-PSDB; AAX06822.
XX
PT   Species-specific test for identifying mammals infected with
PT   Chlamydia pneumoniae - comprises detecting antibodies specific for
PT   outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT   these proteins
```

XX
PS Claim 7; Page 60-62; 115pp; English.
XX
CC This polypeptide comprises the novel 98.4 kDa surface exposed
CC protein Omp10 of the human respiratory pathogen Chlamydia
CC pneumoniae. Its amino acid sequence was deduced from DNA (see
CC AAX06822) isolated from a C. pneumoniae expression library. The
CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
CC (see AAW88417-28), and nucleic acid sequences encoding them (see
CC AAX06816-27). A new species specific test is claimed that is used
CC to identify mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
SQ Sequence 928 AA;

Query Match 8.3%; Score 82; DB 20; Length 928;
Best Local Similarity 21.8%; Pred. No. 17;
Matches 49; Conservative 35; Mismatches 69; Indels 72; Gaps 13;

```
Qy      10 TLNSLFNQIKNQSCGTSTA-SSPCITFRYPVDGCYARAHKMRQILMN--NGYDCE----- 61
           |: :|   :   : |: || :: |  |:        : |  : :| |   | :|
Db      57 tiynltgdvsitnagsptaltasc--fkettgnlsfqghgyqfllqnidaganctftnta 114

Qy      62 -------KQFVY------------GNLKASTGTCCVAWSYHVAILVSYKNASG------ 95
                  | |              | :| ||| | :  :|      :: | :|
Db     115 ankllsfsgfsylsliqttnattgtgaik-stgacsiqsnyscyfgqnfsndnggalqgs 173

Qy      96 ---------------VTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVS--SYANT 137
                         |:|     :|:|:| :|        :| :  ||| |  : ||
Db     174 sislslnpnltfaknkatqk---ggalystggit--------inntlnsasfsentaann 222

Qy     138 AGNVYYRSPSNSYLYDNNLIN-TNCVLTKFSLLSG---CSPSPAP 178
           | :|  : ::|::  |  |:  |  :|  |   |   || : ||
Db     223 ggaiy--teassfissnkaisfinnsvtatsatggaiycsstsap 265
```

Search completed: June 27, 2002, 21:47:22
Job time: 5510 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 20:18:52 ; Search time 38.03 Seconds
(without alignments)
118.820 Million cell updates/sec

Title: US-09-727-769A-6
Perfect score: 991
Sequence: 1 LASVIPDVATLNSLFNQIKN.....FSLLSGCSPSPAPDVSSCGF 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	874	88.2	185	4	US-09-324-910-6	Sequence 6, Appli
2	874	88.2	319	4	US-09-324-910-12	Sequence 12, Appl
3	83	8.4	20	4	US-09-324-910-2	Sequence 2, Appli
4	79	8.0	20	4	US-09-324-910-1	Sequence 1, Appli
5	77.5	7.8	666	4	US-08-982-785A-11	Sequence 11, Appl
6	76.5	7.7	433	4	US-09-104-623A-4	Sequence 4, Appli
7	76.5	7.7	499	2	US-07-952-853-6	Sequence 6, Appli
8	76.5	7.7	499	2	US-08-914-848-6	Sequence 6, Appli
9	76.5	7.7	635	2	US-08-873-479-43	Sequence 43, Appl
10	75	7.6	271	1	US-08-467-155A-10	Sequence 10, Appl
11	75	7.6	271	2	US-08-628-198-10	Sequence 10, Appl
12	75	7.6	271	4	US-09-201-038-10	Sequence 10, Appl
13	75	7.6	271	5	PCT-US96-07343-10	Sequence 10, Appl
14	74.5	7.5	440	1	US-07-930-686-12	Sequence 12, Appl
15	74.5	7.5	440	2	US-08-460-998-12	Sequence 12, Appl
16	73	7.4	659	4	US-08-894-818B-1	Sequence 1, Appli
17	73	7.4	659	4	US-08-894-818B-5	Sequence 5, Appli
18	73	7.4	659	4	US-09-445-472-12	Sequence 12, Appl
19	72	7.3	615	2	US-08-525-742-10	Sequence 10, Appl
20	72	7.3	862	1	US-08-325-267A-4	Sequence 4, Appli
21	72	7.3	894	3	US-08-362-525-22	Sequence 22, Appl
22	72	7.3	894	3	US-08-971-692-15	Sequence 15, Appl
23	72	7.3	1537	1	US-08-325-267A-2	Sequence 2, Appli
24	71.5	7.2	720	3	US-09-257-799-48	Sequence 48, Appl
25	71.5	7.2	720	3	US-08-920-919A-48	Sequence 48, Appl
26	71	7.2	422	2	US-08-712-072C-2	Sequence 2, Appli
27	69	7.0	1627	1	US-07-665-792E-9	Sequence 9, Appli
28	68.5	6.9	294	4	US-09-188-930-294	Sequence 294, App
29	68.5	6.9	544	2	US-08-932-376A-4	Sequence 4, Appli
30	68.5	6.9	572	2	US-08-932-376A-2	Sequence 2, Appli
31	68.5	6.9	583	4	US-08-481-190-19	Sequence 19, Appl
32	68.5	6.9	583	5	PCT-US93-00869-19	Sequence 19, Appl
33	68	6.9	1003	1	US-08-571-758-4	Sequence 4, Appli
34	68	6.9	1003	1	US-08-909-984A-4	Sequence 4, Appli
35	68	6.9	1003	1	US-08-909-983-4	Sequence 4, Appli
36	67.5	6.8	785	3	US-09-265-108-2	Sequence 2, Appli
37	67.5	6.8	785	4	US-09-479-264-2	Sequence 2, Appli
38	67	6.8	647	1	US-08-218-943-1	Sequence 1, Appli
39	67	6.8	661	1	US-08-394-326-2	Sequence 2, Appli
40	67	6.8	661	3	US-09-082-306-2	Sequence 2, Appli
41	67	6.8	1041	1	US-08-220-151-4	Sequence 4, Appli
42	67	6.8	1041	1	US-08-413-118-4	Sequence 4, Appli
43	67	6.8	1041	3	US-08-473-446-4	Sequence 4, Appli
44	67	6.8	1294	2	US-08-819-288-3	Sequence 3, Appli
45	67	6.8	1294	4	US-09-400-348-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-324-910-6
; Sequence 6, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-09-324-910-6

Query Match 88.2%; Score 874; DB 4; Length 185;
Best Local Similarity 87.4%; Pred. No. 1.5e-88;
Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

```
Qy       3 SVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEK 62
           |||||:|||||||  |||||:|||||||||||||||||||||||||||||:| ||||||
Db       3 SVIPDLATLNSLFTQIKNQACGTSTASSPCITFRYPVDGCYARAHKMRQILLNAGYDCEK 62

Qy      63 QFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACV 122
           |||||||:|||||||:| ||||||||:|||||: ||||||||||||||||:||| ||
Db      63 QFVYGNLRASTGTCCVSWVYHVAILVSFKNASGIVEKRIIDPSLFSSGPVTDSAWRAACT 122

Qy     123 NTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSS 182
           |||||||||||||||||||||||| | ||||| :||||||  || ||||||||||  |:|
Db     123 NTSCGSASVSSYANTAGNVYYRSPSGSLLYDNNYVNTNCVLNIFSSLSGCSPSPAPSVAS 182

Qy     183 CGF 185
           |||
Db     183 CGF 185
```

RESULT 2
US-09-324-910-12
; Sequence 12, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro

```
;  APPLICANT: Matsuura, Akira
;  TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
;  TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
;  FILE REFERENCE: Q-54362
;  CURRENT APPLICATION NUMBER: US/09/324,910
;  CURRENT FILING DATE: 1999-06-03
;  EARLIER APPLICATION NUMBER: HEI-10-173940
;  EARLIER FILING DATE: 1998-06-04
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 319
;   TYPE: PRT
;   ORGANISM: Chryseobacterium gleum
US-09-324-910-12


  Query Match             88.2%;  Score 874;  DB 4;  Length 319;
  Best Local Similarity   87.4%;  Pred. No. 3.3e-88;
  Matches  160;  Conservative   10;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       3 SVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEK 62
           |||||:|||||||  |||||:||||||||||||||||||||||||||||||:| ||||||
Db     137 SVIPDLATLNSLFTQIKNQACGTSTASSPCITFRYPVDGCYARAHKMRQILLNAGYDCEK 196

Qy      63 QFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACV 122
           |||||||:|||||||:| ||||||||:|||||: ||||||||||||||||:||| ||
Db     197 QFVYGNLRASTGTCCVSWVYHVAILVSFKNASGIVEKRIIDPSLFSSGPVTDSAWRAACT 256

Qy     123 NTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSS 182
           |||||||||||||||||||||| | ||||| :||||||  || ||||||||| |:|
Db     257 NTSCGSASVSSYANTAGNVYYRSPSGSLLYDNNYVNTNCVLNIFSSLSGCSPSPAPSVAS 316

Qy     183 CGF 185
           |||
Db     317 CGF 319


RESULT   3
US-09-324-910-2
; Sequence 2, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
;  APPLICANT: Yamaguchi, Shotaro
;  APPLICANT: Matsuura, Akira
;  TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
;  TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
;  FILE REFERENCE: Q-54362
;  CURRENT APPLICATION NUMBER: US/09/324,910
;  CURRENT FILING DATE: 1999-06-03
;  EARLIER APPLICATION NUMBER: HEI-10-173940
;  EARLIER FILING DATE: 1998-06-04
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 20
;   TYPE: PRT
;   ORGANISM: Chryseobacterium gleum
US-09-324-910-2


  Query Match              8.4%;  Score 83;  DB 4;  Length 20;
  Best Local Similarity   78.9%;  Pred. No. 0.0081;
  Matches   15;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     145 SPSNSYLYDNNLINTNCVL 163
           ||| | ||||| :||||||
Db       1 SPSGSLLYDNNYVNTNCVL 19


RESULT   4
US-09-324-910-1
; Sequence 1, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
;  APPLICANT: Yamaguchi, Shotaro
;  APPLICANT: Matsuura, Akira
;  TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
;  TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
;  FILE REFERENCE: Q-54362
;  CURRENT APPLICATION NUMBER: US/09/324,910
;  CURRENT FILING DATE: 1999-06-03
;  EARLIER APPLICATION NUMBER: HEI-10-173940
;  EARLIER FILING DATE: 1998-06-04
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 20
;   TYPE: PRT
;   ORGANISM: Chryseobacterium gleum
US-09-324-910-1


  Query Match              8.0%;  Score 79;  DB 4;  Length 20;
  Best Local Similarity   88.9%;  Pred. No. 0.022;
  Matches   16;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SVIPDVATLNSLFNQIKN 20
           |||||:||||||| ||||
Db       3 SVIPDLATLNSLFTQIKN 20


RESULT   5
US-08-982-785A-11
; Sequence 11, Application US/08982785A
; Patent No. 6258929
;  GENERAL INFORMATION:
;    APPLICANT:  Kosik, Kenneth S.
;    APPLICANT:  Zhou, Jianhua
;    TITLE OF INVENTION:  ALARM RELATED PEPTIDES AND
;    TITLE OF INVENTION:  NUCLEIC ACIDS AND DIAGNOSIS USING THEM
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/982,785A
;      FILING DATE:  02-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/031,556
;      FILING DATE:  02-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  05311/018001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  666 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-982-785A-11
```

```
  Query Match             7.8%;  Score 77.5;  DB 4;  Length 666;
  Best Local Similarity   27.5%;  Pred. No. 4.7;
  Matches   38;  Conservative   15;  Mismatches   60;  Indels   25;  Gaps    6;

Qy      41 GCYA-RAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEK 99
           || : :: ||    ||| |||      | | |      | ||   : :| |  | ::|
Db     432 GCESNKSDKM----MNNNYDCPLPEEEINPKGS------GWLYHSDAIRTYLNLMGKSKK 481

Qy     100 RIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSS-YANTAGNVYYRSPSNSYLYDNNLIN 158
                   |  |   |  ||:     :|| :   |      |  : |  :   |
Db     482 -----------DATLEACAGALQNTTASKGLMSSGMSQLIGLKEKGLPQIARLLQSG--N 528

Qy     159 TNCVLTKFSLLSGCSPSP 176
           :: | :  |||| |  |
Db     529 SDVVRSGASLLSNMSKKP 546


RESULT   6
US-09-104-623A-4
; Sequence 4, Application US/09104623A
; Patent No. 6303752
;  GENERAL INFORMATION:
;    APPLICANT:  Olsen, Arne Agerlin
;    APPLICANT:  Fatum, Tine Muxoll
;    APPLICANT:  Deussen, Heinz-Josef
;    APPLICANT:  Roggen, Erwin Ludo
;    TITLE OF INVENTION:  A Modified Polypeptide
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6303752o No. 6303752disk of No. 6303752th America, Inc.
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/104,623A
;      FILING DATE:  25-JUN-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rozek, Carol
;      REGISTRATION NUMBER:  36,993
;      REFERENCE/DOCKET NUMBER:  5256.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  433 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      STRAIN:  Bacillus sp. Y
US-09-104-623A-4


  Query Match             7.7%;  Score 76.5;  DB 4;  Length 433;
  Best Local Similarity   23.0%;  Pred. No. 3.3;
  Matches   45;  Conservative   30;  Mismatches   80;  Indels   41;  Gaps    7;

Qy       9 ATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCY-ARAHKMRQILMNNGYDCEKQFVYG 67
           : ||:||:|  |      | |     :  ||:| | | : :: : : ||  |    |  |
Db     108 SNLNTLFSQAWNAGARIHTNS-----WGAPVNGAYTANSRQVDEYVRNN--DMTVLFAAG 160

Qy      68 NLKASTGTCCVAWSYHVAILV----SYK---------------NASGVTEKRIIDPSLF 107
           |   ::||     :   || |    :|:                :: | |    | | :
Db     161 NEGPNSGTISAPGTAKNAITVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVT 220

Qy     108 SSGPVTDTAWRNACVNTSCGSASVSSYAN----------TAGNVYYRSPSNSYLYDNNLI 157
           : |    :|  :   ::|  :   | ||            ||||   :    :   | |
Db     221 APGTFILSARSSLAPDSSFWANYNSKYAYMGGTSMATPIVAGNV---AQLREHFIKNRGI 277

Qy     158 NTNCVLTKFSLLSGCS 173
               | | :|::| :
Db     278 TPKPSLIKAALIAGAT 293


RESULT   7
US-07-952-853-6
; Sequence 6, Application US/07952853
; Patent No. 5863783
;  GENERAL INFORMATION:
;    APPLICANT:  Van Heuvel, Margaretha
;    APPLICANT:  Bakhuis, Janna G.
;    APPLICANT:  Coutel, Yves
;    APPLICANT:  Harder, Abraham
;    APPLICANT:  De Graaff, Leendert H.
;    APPLICANT:  Flipphi, Michel J. A.
;    APPLICANT:  Van Der Veen, Peter
;    APPLICANT:  Visser, Jacob
;    APPLICANT:  Andreoli, Peter M.
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF DNA
;    TITLE OF INVENTION:  MOLECULES
;    TITLE OF INVENTION:  ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
;    TITLE OF INVENTION:  ORIGIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/952,853
;      FILING DATE:  19921125
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murashige, Kate H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  246152003500
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-813-5600
;      TELEFAX:  415-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-952-853-6


  Query Match             7.7%;  Score 76.5;  DB 2;  Length 499;
  Best Local Similarity   24.7%;  Pred. No. 4.1;
  Matches   59;  Conservative   22;  Mismatches   77;  Indels   81;  Gaps   14;

Qy       1 LASVIPDVATLN-----SLF-----NQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMR 50
           ||| |    |||      :|        |:: ||:|   |         :| ||
Db     120 LASAIGAPVTLNGQKAYGVFMSPGTGYRNNEATGTATGDEP--------EGMYA------ 165

Qy      51 QILMNNGYDCEKQFVYGNLKAS---TGT---------CCVAWSYHVA----ILVSYKN-- 92
           :|    |:    | ||| : |   ||              | |       |:|  :|
```

```
Db    166 -VLDGTHYNDACCFDYGNAETSSTDTGAGHMEAIYLGNSTTWGYGAGDGPWIMVDMENNL 224

Qy     93 ASGVTE-KRIIDPSL---FSSGPVTDTAWRNACVNTSCGSASVSSY---------- 134
Db    225 FSGADEGYNSGDPSISYSFVTAAVKGGADKWAIRGGNAASGSLSTYYSGARPDYSGYNPM 284

Qy    135 -------------ANTAGNVYYRSPSNS-YLYDN--NLINTNCVLTKF---SLLSGCS 173
Db    285 SKEGAIILGIGGDNSNGAQGTFYEGVMTSGYPSDDVENSVQENIVAAKYVSGSLVSGPS 343


RESULT   8
US-08-914-848-6
; Sequence 6, Application US/08914848
; Patent No. 5989887
;  GENERAL INFORMATION:
;    APPLICANT:  Van Heuvel, Margaretha
;    APPLICANT:  Bakhuis, Janna G.
;    APPLICANT:  Coutel, Yves
;    APPLICANT:  Harder, Abraham
;    APPLICANT:  De Graaff, Leendert H.
;    APPLICANT:  Flipphi, Michel J. A.
;    APPLICANT:  Van Der Veen, Peter
;    APPLICANT:  Visser, Jacob
;    APPLICANT:  Andreoli, Peter M.
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF DNA
;    TITLE OF INVENTION:  MOLECULES ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
;    TITLE OF INVENTION:  ORIGIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/914,848
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/952,853
;      FILING DATE:  25-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murashige, Kate H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  246152003500
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-813-5600
;      TELEFAX:  415-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-914-848-6


  Query Match             7.7%;  Score 76.5;  DB 2;  Length 499;
  Best Local Similarity   24.7%;  Pred. No. 4.1;
  Matches   59;  Conservative   22;  Mismatches   77;  Indels   81;  Gaps   14;

Qy      1 LASVIPDVATLN-----SLF-----NQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMR 50
Db    120 LASAIGAPVTLNGQKAYGVFMSPGTGYRNNEATGTATGDEP--------EGMYA------ 165

Qy     51 QILMNNGYDCEKQFVYGNLKAS---TGT---------CCVAWSYHVA----ILVSYKN-- 92
```

```
Db    166 -VLDGTHYNDACCFDYGNAETSSTDTGAGHMEAIYLGNSTTWGYGAGDGPWIMVDMENNL 224

Qy     93 ASGVTE-KRIIDPSL---FSSGPVTDTAWRNACVNTSCGSASVSSY---------- 134
Db    225 FSGADEGYNSGDPSISYSFVTAAVKGGADKWAIRGGNAASGSLSTYYSGARPDYSGYNPM 284

Qy    135 -------------ANTAGNVYYRSPSNS-YLYDN--NLINTNCVLTKF---SLLSGCS 173
Db    285 SKEGAIILGIGGDNSNGAQGTFYEGVMTSGYPSDDVENSVQENIVAAKYVSGSLVSGPS 343


RESULT   9
US-08-873-479-43
; Sequence 43, Application US/08873479
; Patent No. 5891701
;  GENERAL INFORMATION:
;    APPLICANT:  Sloma, Alan
;    APPLICANT:  Lynne, Christianson
;    TITLE OF INVENTION:  Nucleic Acids Encoding A Polypeptide
;    TITLE OF INVENTION:  Having Protease Activity
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5891701o No. 5891701disk of No. 5891701th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/873,479
;      FILING DATE:  12-JUN-1997
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris, Cheryl H
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  5251.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  635 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-873-479-43


  Query Match             7.7%;  Score 76.5;  DB 2;  Length 635;
  Best Local Similarity   23.0%;  Pred. No. 5.7;
  Matches   45;  Conservative   30;  Mismatches   80;  Indels   41;  Gaps    7;

Qy      9 ATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCY-ARAHKMRQILMNNGYDCEKQFVYG 67
Db    310 SNLNTLFSQAWNAGARIHTNS-----WGAPVNGAYTANSRQVDEYVRNN--DMTVLFAAG 362

Qy     68 NLKASTGTCCVAWSYHVAILV----SYK---------------NASGVTEKRIIDPSLF 107
Db    363 NEGPNSGTISAPGTAKNAITVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVT 422

Qy    108 SSGPVTDTAWRNACVNTSCGSASVSSYAN----------TAGNVYYRSPSNSYLYDNNLI 157
Db    423 APGTFILSARSSLAPDSSFWANYNSKYAYMGGTSMATPIVAGNV---AQLREHFIKNRGI 479

Qy    158 NTNCVLTKFSLLSGCS 173
```

```
Db    480 TPKPSLIKAALIAGAT 495


RESULT  10
US-08-467-155A-10
; Sequence 10, Application US/08467155A
; Patent No. 5736377
;  GENERAL INFORMATION:
;    APPLICANT:  Band, Vimla
;    TITLE OF INVENTION:  NES-1 POLYPEPTIDES, DNA, AND RELATED
;    TITLE OF INVENTION:  MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,155A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00398/100001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  271 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-467-155A-10


  Query Match             7.6%;  Score 75;  DB 1;  Length 271;
  Best Local Similarity   23.1%;  Pred. No. 2.5;
  Matches   33;  Conservative   20;  Mismatches   46;  Indels   44;  Gaps    7;

Qy      36 RYPVDGC--YARAHKMRQILMNNGYD-CEKQFVYGNLKASTGTCC-----------VAW 80
           |||:  |  |:: |   |: :|:|| |    |  |   |   |                :
Db      18 RYPIIECKAYSQPH---QVSLNSGYHFCGGSLVNENWVVSAAHCYQSRVEVRLGEHNSST 74

Qy      81 SYHVAILVSYKNASGVTEKRII-------------------DPSLFSSG--PVTDTAWR 118
           | : | |:  :   ::  |:|                    |:  ||   |: :|  :
Db      75 RYPIIIQVTEGSEQFISSSRVIRHPNYSSYNIDNDIMLIKLSKPATLSSTRYPIINTYVQ 134

Qy     119 NACVNTSCGSA----SVSSYANT 137
             : |||  |    :|| : ||
Db     135 PVALPTSCAPAGTMCTVSGWGNT 157


RESULT  11
US-08-628-198-10
; Sequence 10, Application US/08628198
; Patent No. 5843694
;  GENERAL INFORMATION:
;    APPLICANT:  Band, Vimla
;    TITLE OF INVENTION:  NES-1 POLYPEPTIDES, DNA, AND RELATED
;    TITLE OF INVENTION:  MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/628,198
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/467,155
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00398/100002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  271 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-628-198-10


  Query Match             7.6%;  Score 75;  DB 2;  Length 271;
  Best Local Similarity   23.1%;  Pred. No. 2.5;
  Matches   33;  Conservative   20;  Mismatches   46;  Indels   44;  Gaps    7;

Qy      36 RYPVDGC--YARAHKMRQILMNNGYD-CEKQFVYGNLKASTGTCC-----------VAW 80
           |||:  |  |:: |   |: :|:|| |    |  |   |   |                :
Db      18 RYPIIECKAYSQPH---QVSLNSGYHFCGGSLVNENWVVSAAHCYQSRVEVRLGEHNSST 74

Qy      81 SYHVAILVSYKNASGVTEKRII-------------------DPSLFSSG--PVTDTAWR 118
           | : | |:  :   ::  |:|                    |:  ||   |: :|  :
Db      75 RYPIIIQVTEGSEQFISSSRVIRHPNYSSYNIDNDIMLIKLSKPATLSSTRYPIINTYVQ 134

Qy     119 NACVNTSCGSA----SVSSYANT 137
             : |||  |    :|| : ||
Db     135 PVALPTSCAPAGTMCTVSGWGNT 157


RESULT  12
US-09-201-038-10
; Sequence 10, Application US/09201038
; Patent No. 6153387
;  GENERAL INFORMATION:
;    APPLICANT:  Band, Vimla
;    TITLE OF INVENTION:  NES-1 POLYPEPTIDES, DNA, AND RELATED
;    TITLE OF INVENTION:  MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/201,038
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/628,198
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00398/100002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  271 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-201-038-10


  Query Match             7.6%;  Score 75;  DB 4;  Length 271;
  Best Local Similarity   23.1%;  Pred. No. 2.5;
  Matches   33;  Conservative   20;  Mismatches   46;  Indels   44;  Gaps    7;

Qy      36 RYPVDGC--YARAHKMRQILMNNGYD-CEKQFVYGNLKASTGTCC------------VAW 80
           |||:  |  |:: |   |: :|:|| |   |  |  |   |                :
Db      18 RYPIIECKAYSQPH---QVSLNSGYHFCGGSLVNENWVVSAAHCYQSRVEVRLGEHNSST 74

Qy      81 SYHVAILVSYKNASGVTEKRII-------------------DPSLFSSG--PVTDTAWR 118
           | : | |:  :   ::  |:|                    |:  ||   |: :|  :
Db      75 RYPIIIQVTEGSEQFISSSRVIRHPNYSSYNIDNDIMLIKLSKPATLSSTRYPIINTYVQ 134

Qy     119 NACVNTSCGSA----SVSSYANT 137
             : |||  |    :|| : ||
Db     135 PVALPTSCAPAGTMCTVSGWGNT 157


RESULT  13
PCT-US96-07343-10
; Sequence 10, Application PC/TUS9607343
;  GENERAL INFORMATION:
;    APPLICANT:  New England Medical Center Hospitals, Inc.
;    TITLE OF INVENTION:  NES-1 POLYPEPTIDES, DNA, AND RELATED
;    TITLE OF INVENTION:  MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/07343
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/467,155
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00398/100001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  271 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US96-07343-10


  Query Match             7.6%;  Score 75;  DB 5;  Length 271;
  Best Local Similarity   23.1%;  Pred. No. 2.5;
  Matches   33;  Conservative   20;  Mismatches   46;  Indels   44;  Gaps    7;

Qy      36 RYPVDGC--YARAHKMRQILMNNGYD-CEKQFVYGNLKASTGTCC------------VAW 80
           |||:  |  |:: |   |: :|:|| |   |  |  |   |                :
Db      18 RYPIIECKAYSQPH---QVSLNSGYHFCGGSLVNENWVVSAAHCYQSRVEVRLGEHNSST 74

Qy      81 SYHVAILVSYKNASGVTEKRII-------------------DPSLFSSG--PVTDTAWR 118
           | : | |:  :   ::  |:|                    |:  ||   |: :|  :
Db      75 RYPIIIQVTEGSEQFISSSRVIRHPNYSSYNIDNDIMLIKLSKPATLSSTRYPIINTYVQ 134

Qy     119 NACVNTSCGSA----SVSSYANT 137
             : |||  |    :|| : ||
Db     135 PVALPTSCAPAGTMCTVSGWGNT 157


RESULT  14
US-07-930-686-12
; Sequence 12, Application US/07930686
; Patent No. 5525508
;  GENERAL INFORMATION:
;    APPLICANT:  Sharp, Phillip J
;    APPLICANT:  Wagland, Barry M
;    APPLICANT:  Cobon, Gary S
;    TITLE OF INVENTION:  Nematode Vaccine
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley and Lardner
;      STREET:  suite 500, 1800 Diagonal Road
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States of America
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/930,686
;      FILING DATE:  19921006
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PK4486
;      FILING DATE:  06-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU92/00040
;      FILING DATE:  06-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A
;      REGISTRATION NUMBER:  29,768
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-9300
;      TELEFAX:  (703) 683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  12:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  440 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-930-686-12


  Query Match             7.5%;  Score 74.5;  DB 1;  Length 440;
  Best Local Similarity   23.1%;  Pred. No. 5.6;
  Matches   39;  Conservative   22;  Mismatches   59;  Indels   49;  Gaps   10;

Qy      41 GCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVA---------WSYHVAIL--VS 89
           |   :| :| ::    |||||   |  |  |    |            |  :: :|   :
Db      51 GFAPKAARMLKV----GYDCE---VEANTAAYAKECKFEHDPPEQRNYWGQNLWMLGGTN 103

Qy      90 YKNAS----------------GVTEKRIIDPSLFSS--GPVTDTAWRNA----CVNTSCG 127
           |                     || :: |:   :|    |  |  ||:::    |    |
Db     104 YSKTESAKLSVQAWYWELKMFGVPDENILTMEVFDRGVGHYTQVAWQSSDKIGCAVEWCP 163

Qy     128 SAS-VSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPS 175
           : : |:   | |||       | |:||   |   |   :    :||: |
Db     164 TMTLVACEYNPAGN-----RINHYIYD---IGDPCTTDEDCQCTGCTCS 204


RESULT  15
US-08-460-998-12
; Sequence 12, Application US/08460998
; Patent No. 5942413
;  GENERAL INFORMATION:
;    APPLICANT:  Sharp, Phillip J
;    APPLICANT:  Wagland, Barry M
;    APPLICANT:  Cobon, Gary S
;    TITLE OF INVENTION:  Nematode Vaccine
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley and Lardner
;      STREET:  suite 500, 3000 K Street, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  United States of America
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,998
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/930,686
;      FILING DATE:  06-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PK4486
;      FILING DATE:  06-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU92/00040
;      FILING DATE:  06-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A
;      REGISTRATION NUMBER:  29,768
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  440 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-460-998-12


  Query Match             7.5%;  Score 74.5;  DB 2;  Length 440;
  Best Local Similarity   23.1%;  Pred. No. 5.6;
  Matches   39;  Conservative   22;  Mismatches   59;  Indels   49;  Gaps   10;

Qy      41 GCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVA---------WSYHVAIL--VS 89
           |   :| :| ::    |||||   |  |  |    |            |  :: :|   :
Db      51 GFAPKAARMLKV----GYDCE---VEANTAAYAKECKFEHDPPEQRNYWGQNLWMLGGTN 103

Qy      90 YKNAS----------------GVTEKRIIDPSLFSS--GPVTDTAWRNA----CVNTSCG 127
           |                     || :: |:   :|    |  |  ||:::    |    |
Db     104 YSKTESAKLSVQAWYWELKMFGVPDENILTMEVFDRGVGHYTQVAWQSSDKIGCAVEWCP 163

Qy     128 SAS-VSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPS 175
           : : |:   | |||       | |:||   |   |   :    :||: |
Db     164 TMTLVACEYNPAGN-----RINHYIYD---IGDPCTTDEDCQCTGCTCS 204


Search completed: June 27, 2002, 21:48:14
Job time: 5362 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 20:22:32 ; Search time 55.76 Seconds
(without alignments)
318.804 Million cell updates/sec

Title: US-09-727-769A-6
Perfect score: 991
Sequence: 1 LASVIPDVATLNSLFNQIKN.....FSLLSGCSPSPAPDVSSCGF 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	96.5	9.7	534	2	C69146	hypothetical prote
2	92	9.3	551	2	A83232	conserved hypothet
3	92	9.3	2395	1	S50820	surface protein ty
4	91	9.2	594	2	T50765	adhesion of calyx
5	90	9.1	243	2	T28802	hypothetical prote
6	89.5	9.0	927	2	T21772	hypothetical prote
7	89	9.0	1895	2	T15881	hypothetical prote
8	88.5	8.9	283	2	G69212	conserved hypothet
9	88.5	8.9	1251	2	T21389	hypothetical prote
10	85.5	8.6	534	2	T39903	serine-rich protei
11	85.5	8.6	1082	2	S37837	suppressor protein
12	85	8.6	594	2	T50764	adhesion of calyx
13	83.5	8.4	712	2	T33028	hypothetical prote
14	82.5	8.3	5376	2	T42215	zonadhesin - mouse
15	82	8.3	703	2	S67277	probable membrane
16	82	8.3	928	2	E86546	polymorphic outer
17	82	8.3	928	2	B72077	polymorphic membra
18	81.5	8.2	361	2	T19647	hypothetical prote
19	81	8.2	485	2	T49237	hypothetical prote
20	79.5	8.0	207	2	T22436	hypothetical prote
21	79.5	8.0	1894	2	JC4980	plexin 1 precursor
22	79	8.0	221	2	A59040	cysteine proteinas
23	79	8.0	275	2	T43004	hypothetical prote
24	79	8.0	732	2	B69749	hypothetical prote
25	78.5	7.9	725	2	S60712	band-6-protein - h
26	78.5	7.9	1067	2	T31714	probable zinc prot
27	77.5	7.8	422	2	T39306	mitogen-activated
28	77	7.8	814	2	T02195	hypothetical prote
29	77	7.8	1237	2	D71850	probable outer mem
30	77	7.8	2090	2	T30075	hypothetical prote
31	77	7.8	2153	2	T30074	hypothetical prote
32	77	7.8	2718	2	A23475	G surface protein
33	76.5	7.7	499	2	S39113	alpha-L-arabinofur
34	76.5	7.7	505	2	T28859	inward rectifier p
35	76.5	7.7	1234	2	C84465	hypothetical prote
36	76	7.7	735	2	A69146	hypothetical prote
37	76	7.7	934	2	T02742	probable ligand-ga
38	76	7.7	2704	2	S09118	G surface protein
39	75.5	7.6	594	2	T31744	hypothetical prote
40	75.5	7.6	1154	2	A86318	protein F15H18.11
41	75	7.6	443	2	F96990	NADP-specific glut
42	74.5	7.5	439	2	A36385	surface antigen se
43	74.5	7.5	980	2	T30089	probable zinc prot
44	74.5	7.5	1046	2	A26838	prestalk protein p
45	74.5	7.5	1335	2	T17508	glycoprotein Vp260

ALIGNMENTS

RESULT 1
C69146
hypothetical protein MTH359 - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: C69146
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T ; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: C69146
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-534 <MTH>
A;Cross-references: GB:AE000821; GB:AE000666; NID:g2621414; PIDN:AAB84865.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH359

Query Match 9.7%; Score 96.5; DB 2; Length 534;
Best Local Similarity 24.1%; Pred. No. 0.8;
Matches 39; Conservative 23; Mismatches 71; Indels 29; Gaps 4;

```
Qy       5 IPDVATLNSLFNQIKNQS-----CGTSTASSPCITFRYPVD----------------GCY 43
             :|:  |:  | : :||:       | |||     :   |||                :
Db     358 LPNTVTVTRLASSLKNRPENDPYRGESTARYLASSASCPVDSPDIRSLASEITRGLTSTF 417

Qy      44 ARAHK----MRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEK 99
             :||      :|  :  : |   |    | ||  || |      |  :||:   |:|:  :
Db     418 SRAEAVFGWVRDNINYSFYYNTKYGAVGTLKNRTGNCV----DHTHLLVALARAAGIPAR 473

Qy     100 RIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNV 141
             :     |:|| |    |    |  :  :|  :|  |: | |
Db     474 YVHGTCNFTSGNVYGHVWAQLLVGDTWYAADATSSRNSLGVV 515
```

RESULT 2
A83232
conserved hypothetical protein PA3310 [imported] - Pseudomonas aeruginosa (strain PAO
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83232
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L .; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: A83232

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-551 <STO>
A;Cross-references: GB:AE004753; GB:AE004091; NID:g9949433; PIDN:AAG06698.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3310
C;Superfamily: Escherichia coli yidB protein


  Query Match             9.3%;  Score 92;  DB 2;  Length 551;
  Best Local Similarity   21.6%;  Pred. No. 2.1;
  Matches   42;  Conservative   24;  Mismatches   54;  Indels   74;  Gaps    9;

Qy      15 FNQIKNQSCGTSTA-SSPCITFRYPVDGC---YARAHK----------MRQILMNNGYDC 60
           | |::  |||||| | ||:  :|| :      |: |:          :: : : |  ||
Db     274 FTQVR--SCGTSTAVSVPCMFSQYPREDYSDKKAKTHEGLLDILQRAGVQVLWLENNSDC 331

Qy      61 EKQFVYGNLKASTGTC--------------------CVAWSYHVAILVSYKNASGVTEK 99
           :            |||                    |: | | :    :   |: :
Db     332 K-----------GTCLRVPNRDIPKTQPSPFCDGKNCLDESLLVGL---QEYIDGLQDD 376

Qy     100 RII--------DPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYL 151
            ||         |  :   |      ::  |     || |     |              :
Db     377 AIIVLHSDGSHGPEYYERYPKEMERFQPVCRTNQLGSCSKEELVN----------V 422

Qy     152 YDNNLINTNCVLTK 165
           ||| :: |:  |||
Db     423 YDNTILYTDHFLTK 436


RESULT   3
S50820
surface protein type 51B - Paramecium tetraurelia
C;Species: Paramecium tetraurelia
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
C;Accession: S50820
R;Scott, J.; Leeck, C.; Forney, J.
Nucleic Acids Res. 22, 5079-5084, 1994
A;Title: Analysis of the micronuclear B type surface protein gene in Paramecium tetraure
A;Reference number: S50820; MUID:95098630
A;Accession: S50820
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2395 <SCO>
A;Cross-references: EMBL:U07603; NID:g467226; PIDN:AAA81947.1; PID:g467227
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C;Genetics:
A;Genetic code: SGC5
A;Introns: 472/3; 1310/3; 1821/3
C;Superfamily: G surface protein


  Query Match             9.3%;  Score 92;  DB 1;  Length 2395;
  Best Local Similarity   22.4%;  Pred. No. 11;
  Matches   57;  Conservative   22;  Mismatches   76;  Indels  100;  Gaps   12;

Qy      10 TLNSLFNQIKNQSCGTSTASSPCITFRYPVD-----------GCYARAHKMRQILMNNG 57
           |:||      :::|  | | : |       |             || |   :
Db     487 TVNSTNAGCVDKTCENSLAQTIC-------DKDLNNKACIWKGKCYKRECVLASSTTATH 539

Qy      58 YDCEKQFVYGNLKASTGTCCV-------AWSYHVAILVSYKNASGV-------------- 96
           ||:   | |   ::||| ||       | :   |  :  :  |||
Db     540 ADCQTYDV-GCTLSNTGTGCVPLPLKCEAITIEAACNIRLQVTSGVKSYPLCGWNGSSCI 598

Qy      97 ---------TEKRIIDPSLFSSG----------------PVTDTA-------------- 116
                   |     |   : ||                | |  |
Db     599 DKACSTAPKTTATTSDCGTYKSGCVANNPVNGSIQGCQDLPTTCAARKSTENCEITRTGF 658

Qy     117 ----WR---NACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLL 169
               |    :|||  || :|||::   | | :   | :|   |   | |: |:    :
Db     659 PTCLWNSATSACVEKSCSTASVTT---TTGFLTVFSNTNCLAY---LSNSACIAN--NTA 710

Qy     170 SGCSPSPAPDVSSCG 184
           || | |    ||||
Db     711 DGCIPKP----SSCG 721


RESULT   4
T50765
adhesion of calyx edges protein ACE [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T50765
R;Araki, T.; Nakatani-Goto, M.
submitted to the EMBL Data Library, May 1999
A;Description: ADHESION OF CALYX EDGES (ACE), cDNA.
A;Reference number: Z25224
A;Accession: T50765
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-594 <ARA>
A;Cross-references: EMBL:AB027458; PIDN:BAA77837.1
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: ACE
A;Map position: 1


  Query Match             9.2%;  Score 91;  DB 2;  Length 594;
  Best Local Similarity   24.6%;  Pred. No. 2.7;
  Matches   31;  Conservative   14;  Mismatches   45;  Indels   36;  Gaps    5;

Qy      25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMNNGY----DCEKQFVYGNLKAST-- 73
           |:   :| :|| |   |||   |      : ::: :| :     |:|| |:  |  |
Db     442 TNVDDNPSVTFNYFKHPVDLQRCVEAIRLVSKVVTSNRFLNYTQCDKQNVHKMLSLSVKA 501

Qy      74 --------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                              |    | ||   ||       :   ||   |:|| | |
Db     502 NINLRPKQLNDTKSMAQFCKDTVVTIWHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFD 561

Qy     109 SGPVTD 114
           | |:
Db     562 ESPGTN 567


RESULT   5
T28802
hypothetical protein C53B7.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T28802
R;Du, Z.
submitted to the EMBL Data Library, December 1995
A;Description: The sequence of C. elegans cosmid C53B7.
A;Reference number: Z20526
A;Accession: T28802
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-243 <DUZ>
A;Cross-references: EMBL:U42830; PIDN:AAC48274.1; GSPDB:GN00028; CESP:C53B7.3
A;Experimental source: strain Bristol N2; clone C53B7
C;Genetics:
A;Gene: CESP:C53B7.3
A;Map position: X
A;Introns: 8/3; 29/1; 124/1; 164/1


  Query Match             9.1%;  Score 90;  DB 2;  Length 243;
  Best Local Similarity   26.7%;  Pred. No. 1.2;
  Matches   43;  Conservative   14;  Mismatches   64;  Indels   40;  Gaps    9;

Qy      22 SCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKAST--GTCCVA 79
           | | : ::  |::      | |:  :       || ||   |  |||   :|  || |
Db      39 SSGQTCSNGQCMSSLGYNGGLYSNTYN-----NNNQYDMNSQ--YGNQMGTTGSGTYCT- 90
```

```
Qy      80 WSYHVAILVSYKNASGVTEKRIIDPSLFS------------SGPV----TDTAWRNACVN 123
                 ||   || |       | :|            ||      ||      | |
Db      91 --------TSYSCRSGETCMNNRCQSSYSNNMYNSNSQYSNSGSTCRYSTDCYSGQMCSN 142

Qy     124 TSCGSASVSSYANTAGNVYYRSPSNSY-LYDNNL-INTNCV 162
           |    |: |  :  |  | | : :|  |||   :|  ||
Db     143 GVC----VAQYGTSYNNAMYSSSTGNYCTYDNTCGLNQRCV 179


RESULT   6
T21772
hypothetical protein T01D3.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T21772; T24296
R;Steward, C.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19470
A;Accession: T21772
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-927 <WIL>
A;Cross-references: EMBL:Z81527; PIDN:CAB04279.1; GSPDB:GN00023; CESP:T01D3.6
A;Experimental source: clone F35E12
R;Steward, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19870
A;Accession: T24296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-927 <WI2>
A;Cross-references: EMBL:Z81110; PIDN:CAB03262.1; GSPDB:GN00023; CESP:T01D3.6
A;Experimental source: clone T01D3
C;Genetics:
A;Gene: CESP:T01D3.6
A;Map position: 5
A;Introns: 38/2; 87/3; 128/2; 242/2; 374/3; 416/3; 734/3; 842/3; 857/1


  Query Match             9.0%;  Score 89.5;  DB 2;  Length 927;
  Best Local Similarity   20.1%;  Pred. No. 6.2;
  Matches   61;  Conservative   25;  Mismatches   68;  Indels  149;  Gaps   16;

Qy       4 VIPDVATLNSLFNQI--KNQSCGTSTASSPCITFRYPVDGCYARAHKMRQI------LMN 55
           | || |:     |:|  |: ||| :               ||   |::  |
Db     130 VCPDYASGRFCQNEIKCKDNSCGKNA------------DCYVANHQLNCICKPGYTARR 176

Qy      56 NGYDCEKQ------------FVYGNLKAS-TGTCCVAWS--------------------- 81
           || ||: :             |  |:    |||  :|
Db     177 NGRDCDMKVQQACMSGDPHYVTYDGLRFDYQGTCPYVFSQPCTTLPAPYLWYSVRAKNEL 236

Qy      82 ----YHVA----ILVSYKN------------ASGVTEKRIIDPSLFS-----------S 109
               ||:    : |   |              :||   ::: |  |            |
Db     237 PGKGYHISQVSEVEVDLHNLTIHVDGRSKTALVNGV---QVLTPWYFPNKNTWTVRVRFS 293

Qy     110 GP-----------VTDTAWRNACV-----------NTSCGSA------SVSSYANTAGNV 141
           |            || | : : ||            | || |       :    |  |:|
Db     294 GSTFTIENDQGVVVTFTTYNSLCVQVPDIPEFNGATTLCGLAGNIDGKKLDDVVNKNGSV 353

Qy     142 Y----YRSPSN-----------SYLYDNNLI---------------NTNCVLTKFSLLS 170
           |    | | |           ::: |  ||               |||| |  ||
Db     354 LAIKSSRQPENNNHADFMKTEDTWITDKFLILRPGQENCINGQTLDNNTNCVSTSISLAQ 413

Qy     171 GCS 173
            |:
Db     414 SCA 416


RESULT   7
T15881
hypothetical protein D1044.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15881
R;Pauley, A.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of C. elegans cosmid D1044.
A;Reference number: Z18423
A;Accession: T15881
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1895 <PAU>
A;Cross-references: EMBL:U00065; NID:g495681; PID:g495684; PIDN:AAA50735.1; CESP:D104
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:D1044.3
A;Introns: 35/2; 63/2; 150/1; 191/3; 207/2; 405/3; 551/3; 629/1; 764/1; 933/1; 959/2;


  Query Match             9.0%;  Score 89;  DB 2;  Length 1895;
  Best Local Similarity   24.1%;  Pred. No. 15;
  Matches   46;  Conservative   27;  Mismatches   88;  Indels   30;  Gaps    9;

Qy      20 NQSC----------GTSTASSPCITFRYPVDGCYARAHKM-RQILMNNGYDCEKQF--VY 66
           || |           || : :| |||   |   |   :  :   : ||    |   :  ||
Db    1112 NQQCVISNGLNCQLGTVSYNSQCITLASPGQNCQTSSQCIDNSVCMNQMCTCNNNYRLVY 1171

Qy      67 GNLKASTGTCCVAWSYHV---AILVSYKNASGVTEKRIIDPSLFSSGPVT----DTAWRN 119
           |     | : |      |    :|:|    : :  :: :  :: :||        ||
Db    1172 GYCVPITSSICQQTQTLVNNQCVLLSIVGETCIANQQCVGGAMCNSGTCQCTNGATAMYG 1231

Qy     120 ACVNTSCGSASVSSYANTAGNVYYR-----SPSNSYLYDNNLINTN--CVLTKFSLLSGC 172
           |:::|  | : |:  :  |  |       | | |    || : ||  || |  |:   |
Db    1232 YCISSSSSSCN-SNQVSINGMCYNTVQVGGSCSFSQQCLNNAVCTNNICVSTFCSV--SC 1288

Qy     173 SPSPAPDVSSC 183
           | :     : |
Db    1289 STNQVCISNQC 1299


RESULT   8
G69212
conserved hypothetical protein MTH845 - Methanobacterium thermoautotrophicum (strain
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: G69212
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: G69212
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-283 <MTH>
A;Cross-references: GB:AE000861; GB:AE000666; NID:g2621930; PIDN:AAB85343.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH845


  Query Match             8.9%;  Score 88.5;  DB 2;  Length 283;
  Best Local Similarity   24.6%;  Pred. No. 2;
  Matches   29;  Conservative   17;  Mismatches   61;  Indels   11;  Gaps    2;

Qy      41 GCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGT-------CCVAWSYHVAILVSYKNA 93
           || :   |  :|         :  | |   | : ||       ||     |  :||:   |
Db     161 GCNSTYDKAVRIFNWVRDSIDYSFYYNTRKGAVGTLHSGSANCC----DHTHLLVALARA 216

Qy      94 SGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYL 151
           ||:  : :    :| ||      |    ||     |  :|::|  | |     |::::
Db     217 SGIPARYMHGNCVFRSGNTYGHVWGQLYVNGRWYDADATSFSNALGTVNNWDRSSAFI 274


RESULT   9
```

```
T21389
hypothetical protein F26C11.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21389
R;Matthews, P.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19416
A;Accession: T21389
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1251 <WIL>
A;Cross-references: EMBL:Z47072; PIDN:CAA87369.1; GSPDB:GN00020; CESP:F26C11.3
A;Experimental source: clone F26C11
C;Genetics:
A;Gene: CESP:F26C11.3
A;Map position: 2
A;Introns: 24/1; 111/1; 208/1; 272/1; 380/1; 394/3; 485/3; 586/3; 630/3; 669/3; 713/3; 7

  Query Match             8.9%;  Score 88.5;  DB 2;  Length 1251;
  Best Local Similarity   23.7%;  Pred. No. 11;
  Matches   36;  Conservative   24;  Mismatches   55;  Indels   37;  Gaps    6;

Qy      21 QSC-GTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVA 79
           | |  |||  :|  || :|  |   |     :|::: 
Db     986 QGCKATSTTQTP-TTFNWPTGGT-TRTLPSGEIILSES---------------------- 1021

Qy      80 WSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAG 139
             |::||| : |  : | :||  ::   | |:    |  || |: |  |   | |
Db    1022 -------LIAYKNCTTVLMQLIYNPSKNTTRTET-TSDAEGCKATSSGTTSTMSPGTTGG 1073

Qy     140 NVYYRSPSNSYLYDNNLINTNCVLTKFSLLSG 171
             |: :::   |:: : |    | |:  :|
Db    1074 TTVSRTTNSNNPIDSSTLET----TTFAWPTG 1101


RESULT  10
T39903
serine-rich protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C;Accession: T39903
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z21889
A;Accession: T39903
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-534 <LYN>
A;Cross-references: EMBL:AL033534; PIDN:CAA22127.1; GSPDB:GN00067; SPDB:SPBC215.13
A;Experimental source: strain 972h-; cosmid c215
C;Genetics:
A;Gene: SPDB:SPBC215.13
A;Map position: 2
C;Superfamily: pig submaxillary mucin

  Query Match             8.6%;  Score 85.5;  DB 2;  Length 534;
  Best Local Similarity   24.2%;  Pred. No. 7.4;
  Matches   45;  Conservative   32;  Mismatches   86;  Indels   23;  Gaps    5;

Qy       6 PDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFV 65
           |  ::  |:|::    |  |:: |||  ::  |            |:  :  :
Db      78 PTSSSEPSIFSESATPS-ETNSYSSPVSSYSDPATS--------QLPSSTSFFSPTSSE 127

Qy      66 YGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTS 125
           |     : :     |  ||| |    |   |  |   || || |:| : : :   :||
Db     128 YTPSSTESSSLLDPSSVSSAILPS----STSVEVSISSSSLSSSDPLTSSTFSSLSSSTS 183

Qy     126 CGSASVSS-YANTAGNVYYRSPSNSYLYDNNLI--------NTNCVLTKFSLLSGCSPSP 176
             ||||  ::|  :    | |:|||  ::::        :::  ||  || :   ||
Db     184 SSQPSVSSTSSSTFSSAAPTSTSSSYLSSSSVVSSSSSPSSSSSSTLTSSSLSTSSIPST 243

Qy     177 APDVSS 182
             :  ||
Db     244 SSSSSS 249


RESULT  11
S37837
suppressor protein SPT23 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKL020c
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 20-Sep-1999
C;Accession: S37837; S40696
R;Rieger, M.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37832
A;Accession: S37837
A;Molecule type: DNA
A;Residues: 1-1082 <RIE>
A;Cross-references: EMBL:Z28020; NID:g486010; PIDN:CAA81855.1; PID:g486011; MIPS:YKL0
A;Experimental source: strain S288C
R;Burkett, T.J.; Garfinkel, D.J.
Yeast 10, 81-92, 1994
A;Title: Molecular characterization of the SPT23 gene: a dosage-dependent suppressor
A;Reference number: S40695; MUID:94262317
A;Accession: S40696
A;Molecule type: DNA
A;Residues: 136-714,'P',716-737,'ERKIYLVLLATYGVY' <BUR>
A;Cross-references: EMBL:L24760; NID:g404815; PIDN:AAA20575.1; PID:g404817
C;Genetics:
A;Gene: SGD:SPT23
A;Cross-references: SGD:S0001503; MIPS:YKL020c
A;Map position: 11L
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
F;742-774/Domain: ankyrin repeat homology <ANl>

  Query Match             8.6%;  Score 85.5;  DB 2;  Length 1082;
  Best Local Similarity   21.7%;  Pred. No. 17;
  Matches   40;  Conservative   27;  Mismatches   54;  Indels   63;  Gaps    8;

Qy      27 TASSPCITFRYPVD-GCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVA 85
           : :| || |  |    || | ||       ||:                        |
Db     259 SGNSNCINFDLPTRIVCYCRHHK-----ATNGF------------------------VV 288

Qy      86 ILVSYKNASGVTEKRIIDPSLF-----SSGPVTDTAWRNACVNTSCGSASVSSYANTAGN 140
           : :   :   :  | | || :      :|   | |:  || |:    :    |::: :
Db     289 LFLLRDHNGDILAKTITDPIMIMDKKNASNTTTPTSTSNAQVSPM--TNDTRSFSSPQSD 346

Qy     141 VYYRS----PSNSYLYDNNLINTNCVL-----------------TKFSLLSGCSPSPAP 178
           : : |    ||||    | :|:|||:|                  |  ::::     |:|
Db     347 LNFPSEFPLPSNS---KNFVISTNCMLDSNCNNNNNDNDNKNNIKTNTAMMNNNRHFPSP 403

Qy     179 DVSS 182
           : ||
Db     404 NSSS 407


RESULT  12
T50764
adhesion of calyx edges protein ACE [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T50764
R;Araki, T.; Nakatani-Goto, M.
submitted to the EMBL Data Library, May 1999
A;Description: Arabidopsis ADHESION OF CALYX EDGES (ACE), genomic.
A;Reference number: Z25223
A;Accession: T50764
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-594 <ARA>
A;Cross-references: EMBL:AB027507; PIDN:BAA77842.1
```

```
A;Experimental source: strain Nossen
C;Genetics:
A;Gene: ACE
A;Introns: 25/1; 155/2; 259/1; 387/3; 577/2


  Query Match             8.6%;  Score 85;  DB 2;  Length 594;
  Best Local Similarity   24.6%;  Pred. No. 9.3;
  Matches   31;  Conservative   13;  Mismatches   46;  Indels   36;  Gaps    5;

Qy      25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMN----NGYDCEKQFVYGNLKAST-- 73
           |:   :| :|| |   |||   |      : ::: :    |   |:|| |:  |  |
Db     442 TNVDDNPSVTFNYFKHPVDLQRCVEAIRLVSKVVTSKRFLNYTQCDKQNVHKMLSLSVKA 501

Qy      74 -------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                               |   | ||  ||       :   ||   |:|| | |
Db     502 NINLRPKQLNDTKSMAQFCKDTVVTIWHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFD 561

Qy     109 SGPVTD 114
           | |:
Db     562 ESPGTN 567


RESULT  13
T33028
hypothetical protein T09A12.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33028
R;Hawkins, J.; Fulton, B.; Gillam, B.
submitted to the EMBL Data Library, February 1998
A;Description: The sequence of C. elegans cosmid T09A12.
A;Reference number: Z21265
A;Accession: T33028
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-712 <HAW>
A;Cross-references: EMBL:AF047660; PIDN:AAC04434.1; GSPDB:GN00022; CESP:T09A12.4
A;Experimental source: strain Bristol N2; clone T09A12
C;Genetics:
A;Gene: CESP:T09A12.4
A;Map position: 4
A;Introns: 32/1; 158/1; 220/1; 244/3; 295/1; 319/1; 436/3; 543/2; 634/3


  Query Match             8.4%;  Score 83.5;  DB 2;  Length 712;
  Best Local Similarity   21.5%;  Pred. No. 15;
  Matches   49;  Conservative   20;  Mismatches   66;  Indels   93;  Gaps   11;

Qy      23 CGTSTASSPCITFRYPVDGCYA-RAHKMRQILMNNGYDCEKQFVYGNLK-ASTGTCCVAW 80
           ||| :         : || | |  |   | :::|  | | |      :|  | |  | |
Db     251 CGTDSTG-----IHFGVDACAACSAFFRRTVVLNKDYSCNKGGKCTVVKDGSAGQKCRAC 305

Qy      81 SYHVAILVSY-------------KNASGVTEKRIIDPSL---------------FSSGP- 111
           :   |                  | ::||  || : |                  ||||
Db     306 RFRKCISSGMDKNSVQHRRDAIGKYSAGV--KRELSPDAEFEPSAKVSTVSEPSTSSGPS 363

Qy     112 ------------------VTDTAWRNACVNTSCGSASVSS----------------YANT 137
                            |   : |:|::||  ||                    ||
Db     364 GGFNQNVSSPAGIPRVPSTLRTTQASTCMNSACGQKSVLHELICRQNFLTEQRQLFYAGC 423

Qy     138 AGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCGF 185
           |: ::| ||:    :|  :                |   | ||| |
Db     424 LGD-WFRKPSS---IENQTL----------------SELTDFSSCMF 450


RESULT  14
T42215
zonadhesin - mouse
N;Alternate names: sperm-specific membrane protein
C;Species: Mus musculus (house mouse)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T42215
R;Gao, Z.; Garbers, D.L.
J. Biol. Chem. 273, 3415-3421, 1998
A;Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane
A;Reference number: Z22080; MUID:98123114
A;Accession: T42215
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5376 <GAO>
A;Cross-references: EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC26680.1
C;Genetics:
A;Gene: Zan
A;Map position: 5
C;Function:
A;Description: functions in multiple cell adhesion processes
A;Note: found exclusively on the apical region of the sperm head
C;Keywords: cell adhesion


  Query Match             8.3%;  Score 82.5;  DB 2;  Length 5376;
  Best Local Similarity   24.9%;  Pred. No. 1.9e+02;
  Matches   43;  Conservative   12;  Mismatches   53;  Indels   65;  Gaps   10;

Qy      24 GTS-TASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSY 82
           ||| | | |       :||                  ||  :|  | |    : |
Db    4530 GTSPEAPSTC------EEGCV---------------CEPDYVLSNDKCVPSSEC----- 4562

Qy      83 HVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAW--RNACVNTSCGSASVSSYANTAGN 140
                 |:| ||    :|  |           |  |    | :|   :|  :
Db    4563 ------GCKDAHGV----LIPES---------KTWVSRGCTKNCTCKGGTVQCHD----- 4598

Qy     141 VYYRSPSNSYLYDNNLINTNCV-----LTKFSLLSGCSPSPAPDVSS----CG 184
           :  |: |   |||  |:|||         || : | ||  |  |     ||
Db    4599 --FSCPTGSRCLDNNEGNSNCVTYALKCPAHSLYTNCLPSCLPSCSDPEGLCG 4649


RESULT  15
S67277
probable membrane protein YOR365c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein O6661
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 20-Jun-2000
C;Accession: S67277
R;Delius, H.; Hebling, U.; Hofmann, B.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67261
A;Accession: S67277
A;Molecule type: DNA
A;Residues: 1-703 <DEL>
A;Cross-references: EMBL:Z75272; NID:g1420789; PIDN:CAA99695.1; PID:g1420791; GSPDB:G
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YOR365c
A;Map position: 15R
C;Superfamily: probable membrane protein YAL053w
C;Keywords: transmembrane protein
F;2-18/Domain: transmembrane #status predicted <TM1>
F;208-224/Domain: transmembrane #status predicted <TM2>
F;248-264/Domain: transmembrane #status predicted <TM3>
F;379-395/Domain: transmembrane #status predicted <TM4>
F;431-447/Domain: transmembrane #status predicted <TM5>
F;454-470/Domain: transmembrane #status predicted <TM6>
F;546-562/Domain: transmembrane #status predicted <TM7>
F;575-591/Domain: transmembrane #status predicted <TM8>
F;599-615/Domain: transmembrane #status predicted <TM9>
F;616-632/Domain: transmembrane #status predicted <TM10>


  Query Match             8.3%;  Score 82;  DB 2;  Length 703;
  Best Local Similarity   19.0%;  Pred. No. 21;
  Matches   39;  Conservative   34;  Mismatches   66;  Indels   66;  Gaps    6;

Qy      31 PCITFRYP-------------VDGCYARAHKMRQILMNNGYDCEKQFVYGN--------- 68
           | ||:  |             :|  : :     |::::||   : :::  |
```

```
Db    157 PSITYNIPDLDAYIVVTAYSTIDKEFTKPLACVQVMLSNGRTVQTEYLSWNLVILTISGI 216

Qy     69 -------LKASTGTCCVAWSYHVAILVSYKN------------------------ASGV 96
                |:  | |     || :::|: ::|                          : |:
Db    217 MFSVVYSLQGYTVTSTRLASYSISLLLYFQNLAILAMISVSFLPPIVAAWTQNFQWSMGI 276

Qy     97 TE----KRIIDPSLFSSGPVTDTAWRNACV------NTSCGSASVSSYANTAGNVYYRSP 146
             :|: |  : ::       :||  |         |  |  :|: :|  |     |
Db    277 IRINFMQRLFDWYVVATSGSPTVVYRNKEVLSISVQKRSLNSKIISASSNLNG--IESSQ 334

Qy    147 SNSYLYDNNLINTNCVLTKFSLLSG 171
           |  || :|| |:|  |:|  :| |
Db    335 KNDLLYTSNLRNSNDYLSKILVLRG 359


Search completed: June 27, 2002, 21:49:25
Job time: 5213 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:48:17 ; Search time 39.56 Seconds
(without alignments)
181.070 Million cell updates/sec

Title: US-09-727-769A-6
Perfect score: 991
Sequence: 1 LASVIPDVATLNSLFNQIKN.....FSLLSGCSPSPAPDVSSCGF 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	89	9.0	1895	1	YLK3_CAEEL	P41951 caenorhabdi
2	88.5	8.9	1251	1	YQU3_CAEEL	Q09550 caenorhabdi
3	86	8.7	344	1	CYS5_DICDI	P54640 dictyosteli
4	85.5	8.6	1082	1	SP23_YEAST	P35210 saccharomyc
5	82.5	8.3	5376	1	ZAN_MOUSE	O88799 mus musculu
6	82	8.3	928	1	PMP9_CHLPN	Q9z398 chlamydia p
7	82	8.3	2715	1	G156_PARPR	P13837 paramecium
8	79	8.0	221	1	GPI_ZINOF	P82473 zingiber of
9	77.5	7.8	422	1	SPM1_SCHPO	Q92398 schizosacch
10	77.5	7.8	728	1	PKP1_MOUSE	P97350 mus musculu
11	76.5	7.7	499	1	ABFB_ASPNG	P42255 aspergillus
12	76.5	7.7	505	1	IRK4_CAEEL	P52192 caenorhabdi
13	76.5	7.7	747	1	PKP1_HUMAN	Q13835 homo sapien
14	76	7.7	2704	1	G168_PARPR	P17053 paramecium
15	75.5	7.6	603	1	CFAI_MOUSE	Q61129 mus musculu
16	74.5	7.5	1046	1	PSTA_DICDI	P11976 dictyosteli
17	74.5	7.5	1047	1	HIRA_DROME	O17468 drosophila
18	74	7.5	364	1	YBBB_ECOLI	P33667 escherichia
19	74	7.5	1672	1	PMPB_CHLMU	Q9pjy2 chlamydia m
20	74	7.5	3305	1	APLP_MANSE	Q25490 manduca sex
21	73.5	7.4	802	1	CSD2_ECOLI	P53513 escherichia
22	73.5	7.4	1679	1	YMF9_YEAST	Q04958 saccharomyc
23	73	7.4	3726	1	TRX_DROME	P20659 drosophila
24	72.5	7.3	402	1	OPS4_CANAL	P46596 candida alb
25	72	7.3	313	1	YM58_YEAST	Q03695 saccharomyc
26	72	7.3	533	1	CAG2_MOUSE	Q09200 mus musculu
27	72	7.3	730	1	GLN3_YEAST	P18494 saccharomyc
28	72	7.3	1110	1	VGLM_INSV	Q01260 impatiens n
29	72	7.3	1136	1	C4BA_BACTI	P05519 bacillus th
30	72	7.3	1376	1	VGL2_CVMJC	Q02385 murine coro
31	72	7.3	1537	1	FLO1_YEAST	P32768 saccharomyc
32	72	7.3	4548	1	APOA_HUMAN	P08519 homo sapien
33	71.5	7.2	666	1	YEA7_YEAST	P40002 saccharomyc
34	71.5	7.2	720	1	TGLX_HUMAN	O43548 homo sapien
35	71.5	7.2	1162	1	VGL2_IBVB	P11223 avian infec
36	71	7.2	217	1	CATS_BOVIN	P25326 bos taurus
37	71	7.2	406	1	LMP1_MOUSE	P11438 mus musculu
38	71	7.2	470	1	RAN1_SCHPO	P08092 schizosacch
39	71	7.2	741	1	HOX1_HALRO	P28468 halocynthia
40	70.5	7.1	201	1	YADL_ECOLI	P37017 escherichia
41	70.5	7.1	574	1	YB2A_SCHPO	P87311 schizosacch
42	70.5	7.1	690	1	TRFE_ORYLA	P79819 oryzias lat
43	70.5	7.1	725	1	NCA2_MOUSE	P13594 mus musculu
44	70.5	7.1	727	1	PKP1_BOVIN	Q28161 bos taurus
45	70.5	7.1	761	1	NCA2_HUMAN	P13592 homo sapien

ALIGNMENTS

RESULT 1
YLK3_CAEEL
ID YLK3_CAEEL STANDARD; PRT; 1895 AA.
AC P41951;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative serine/threonine-protein kinase D1044.3 in chromosome III
DE (EC 2.7.1.-).
GN D1044.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Pauley A., Waterston R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STRONG, TO ZC84.1.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U00065; AAA50735.1; -.
DR HSSP; Q63450; 1A06.
DR WormPep; D1044.3; CE01206.
DR InterPro; IPR002899; EB.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF01683; EB; 12.
DR Pfam; PF00069; pkinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00289; WR1; 12.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW ATP-binding.
FT DOMAIN 431 703 PROTEIN KINASE.
FT NP_BIND 437 445 ATP (BY SIMILARITY).
FT BINDING 461 461 ATP (BY SIMILARITY).
FT ACT_SITE 569 569 BY SIMILARITY.
SQ SEQUENCE 1895 AA; 208393 MW; F23C9F7881353AD6 CRC64;

Query Match 9.0%; Score 89; DB 1; Length 1895;
Best Local Similarity 24.1%; Pred. No. 2.5;
Matches 46; Conservative 27; Mismatches 88; Indels 30; Gaps 9;

```
Qy      20 NQSC----------GTSTASSPCITFRYPVDGCYARAHKM-RQILMNNGYDCEKQF--VY 66
           || |          || : :| |||    |   |   :   :  : ||    |   :  ||
Db    1112 NQQCVISNGLNCQLGTVSYNSQCITLASPGQNCQTSSQCIDNSVCMNQMCTCNNNYRLVY 1171

Qy      67 GNLKASTGTCCVAWSYHV---AILVSYKNASGVTEKRIIDPSLFSSGPVT----DTAWRN 119
           |     | : |       |    :|:|    : :  :: :  :: :||        ||
Db    1172 GYCVPITSSICQQTQTLVNNQCVLLSIVGETCIANQQCVGGAMCNSGTCQCTNGATAMYG 1231

Qy     120 ACVNTSCGSASVSSYANTAGNVYYR-----SPSNSYLYDNNLINTN--CVLTKFSLLSGC 172
           |:::|  | : |:  :  |  |       | | |    || : ||  || |  |:   |
Db    1232 YCISSSSSSCN-SNQVSINGMCYNTVQVGGSCSFSQQCLNNAVCTNNICVSTFCSV--SC 1288

Qy     173 SPSPAPDVSSC 183
           | :     : |
Db    1289 STNQVCISNQC 1299


RESULT   2
YQU3_CAEEL
ID   YQU3_CAEEL     STANDARD;      PRT;  1251 AA.
AC   Q09550;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Hypothetical 133.5 kDa protein F26C11.3 in chromosome II.
GN   F26C11.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Matthews P.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z47072; CAA87369.1; -.
DR   WormPep; F26C11.3; CE01561.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 1.
DR   SMART; SM00032; CCP; 1.
KW   Hypothetical protein.
FT   DOMAIN      120    414       SER/THR-RICH.
SQ   SEQUENCE   1251 AA;  133498 MW;  2B959ECA03B9954A CRC64;


  Query Match             8.9%;  Score 88.5;  DB 1;  Length 1251;
  Best Local Similarity   23.7%;  Pred. No. 1.7;
  Matches   36;  Conservative   24;  Mismatches   55;  Indels   37;  Gaps    6;

Qy      21 QSC-GTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVA 79
           | |  |||  :|  || :|  |   |     :|:::
Db     986 QGCKATSTTQTP-TTFNWPTGGT-TRTLPSGEIILSES---------------------- 1021

Qy      80 WSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAG 139
              |::||| : |  : | :||  ::   | |:    |  || |: |  |   | | |
Db    1022 -------LIAYKNCTTVLMQLIYNPSKNTTRTET-TSDAEGCKATSSGTTSTMSPGTTGG 1073

Qy     140 NVYYRSPSNSYLYDNNLINTNCVLTKFSLLSG 171
             |: :::   |:: : |   | |:  :|
Db    1074 TTVSRTTNSNNPIDSSTLET----TTFAWPTG 1101


RESULT   3
CYS5_DICDI
ID   CYS5_DICDI     STANDARD;      PRT;   344 AA.
AC   P54640;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Cysteine proteinase 5 precursor (EC 3.4.22.-).
GN   CPRE OR CP5.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=96081966; PubMed=7499424;
RA   Souza G.M., Hirai J., Mehta D.P., Freeze H.H.;
RT   "Identification of two novel Dictyostelium discoideum cysteine
RT   proteinases that carry N-acetylglucosamine-1-P-modification.";
RL   J. Biol. Chem. 270:28938-28945(1995).
CC   -!- SUBCELLULAR LOCATION: Lysosomal.
CC   -!- DEVELOPMENTAL STAGE: PRESENT IN THE VEGETATIVE PHASE AND DECREASES
CC       WITH THE START DEVELOPMENT, REAPPEARS IN LOW LEVELS WHEN THE
CC       FRUITING BODY IS FORMED.
CC   -!- PTM: PHOSPHOGLYCOSYLATED, CONTAINS GLCNAC-ALPHA-1-P-SER RESIDUES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC       PAPAIN FAMILY OF THIOL PROTEASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L36205; AAA92018.1; -.
DR   HSSP; P07711; 1CJL.
DR   MEROPS; C01.081; -.
DR   DictyDb; DD01061; cprE.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Thiolprot_act_site.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW   Hydrolase; Thiol protease; Lysosome; Zymogen; Glycoprotein;
KW   Phosphorylation; Signal.
FT   SIGNAL       1     17       POTENTIAL.
FT   PROPEP      18    111       ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN      112    344       CYSTEINE PROTEINASE 5.
FT   DOMAIN     196    340       SER-RICH.
FT   ACT_SITE   136    136       BY SIMILARITY.
FT   ACT_SITE   272    272       BY SIMILARITY.
FT   ACT_SITE   311    311       BY SIMILARITY.
FT   DISULFID   133    174       BY SIMILARITY.
FT   DISULFID   167    207       BY SIMILARITY.
FT   DISULFID   265    333       BY SIMILARITY.
FT   CARBOHYD   110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   297    297       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   344 AA;  37212 MW;  82F3F35B8C147BA8 CRC64;


  Query Match             8.7%;  Score 86;  DB 1;  Length 344;
  Best Local Similarity   25.4%;  Pred. No. 0.66;
  Matches   49;  Conservative   19;  Mismatches   73;  Indels   52;  Gaps    9;

Qy      26 STASSPCITFRYPVDG---CYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSY 82
           || :| |       ||    ||       |: ||| | |  : |   ||  | |
Db     168 STENSGC-------DGGLMTYA----FEYIINNNGIDTESSYPY---KAENGKCEYKSEN 213

Qy      83 HVAILVSYKNASGVTEKRI-----------IDPS-----LFSSGPVTDTAWRNAC---- 121
             | | |||  :  :|  :            || |     |::||      :   |
Db     214 SGATLSSYKTVTAGSESSLESAVNVNPVSVAIDASHQSFQLYTSG----IYYEPECSSEN 269
```

```
Qy    122 ---------VNTSCGSASVSSYANTAGNVYYYRSPSNSYLYDNNLINTNCVLTKFSLLSGC 172
                   :  ||:|  |   ::||:   | || |    |   |:  :  : |:|
Db    270 LDHGVLAVGYGSGSGSSSGQSSGQSSGNL-SASSSNEYWIVKNSWGTSWGIEGYILMSRN 328

Qy    173 SPSPAPDVSSCGF 185
          :    ||  |
Db    329 RDNNCGIASSASF 341
```

RESULT 4
SP23_YEAST
ID SP23_YEAST STANDARD; PRT; 1082 AA.
AC P35210;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SPT23 protein.
GN SPT23 OR YKL020C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieger M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 136-752 FROM N.A.
RX MEDLINE=94262317; PubMed=8203154;
RA Burkett T.J., Garfinkel D.J.;
RT "Molecular characterization of the SPT23 gene: a dosage-dependent
RT suppressor of Ty-induced promoter mutations from Saccharomyces
RT cerevisiae.";
RL Yeast 10:81-92(1994).
CC -!- FUNCTION: DOSAGE-DEPENDENT SUPPRESSOR OF TY-INDUCED PROMOTER
CC MUTATIONS. MAY EXERT ITS SUPPRESSION EFFECT THROUGH PROTEIN-
CC PROTEIN INTERACTIONS SINCE DOES NOT PRESENT ANY OF THE MOTIFS
CC GENERALLY FOUND IN TRANSCRIPTIONAL ACTIVATORS OR DNA BINDING
CC PROTEINS.
CC -!- SIMILARITY: CONTAINS 2 ANK REPEATS.
CC -!- SIMILARITY: TO YEAST MGA2.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 738
CC ONWARD AND IS SHORTER (752 AA) DUE TO A FRAMESHIFT.
CC ---

CC ---
DR EMBL; Z28020; CAA81855.1; -.
DR EMBL; L24760; AAA20575.1; ALT_FRAME.
DR PIR; S37837; S37837.
DR SGD; S0001503; SPT23.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF00023; ank; 2.
DR Pfam; PF01833; TIG; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00429; IPT; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
KW ANK repeat; Repeat.
FT REPEAT 709 738 ANK 1.
FT REPEAT 742 771 ANK 2.
FT CONFLICT 715 715 H -> P (IN REF. 2).
SQ SEQUENCE 1082 AA; 121337 MW; 620C688ECC0ECBD7 CRC64;

Query Match 8.6%; Score 85.5; DB 1; Length 1082;
Best Local Similarity 21.7%; Pred. No. 2.7;
Matches 40; Conservative 27; Mismatches 54; Indels 63; Gaps 8;

```
Qy     27 TASSPCITFRYPVD-GCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVA 85
          : :| || |  |    || | ||       ||:                        |
Db    259 SGNSNCINFDLPTRIVCYCRHHK-----ATNGF-------------------------VV 288

Qy     86 ILVSYKNASGVTEKRIIDPSLF-----SSGPVTDTAWRNACVNTSCGSASVSSYANTAGN 140
          : :   :   :  | | || :      :|   | |:  || |:    :    |:::   :
Db    289 LFLLRDHNGDILAKTITDPIMIMDKKNASNTTTPTSTSNAQVSPM--TNDTRSFSSPQSD 346

Qy    141 VYYRS----PSNSYLYDNNLINTNCVL-----------------TKFSLLSGCSPSPAP 178
          : : |    ||||    | :|:|||:|                  |  ::::     |:|
Db    347 LNFPSEFPLPSNS---KNFVISTNCMLDSNCNNNNNDNDNKNNIKTNTAMMNNNRHFPSP 403

Qy    179 DVSS 182
          : ||
Db    404 NSSS 407
```

RESULT 5
ZAN_MOUSE
ID ZAN_MOUSE STANDARD; PRT; 5376 AA.
AC O88799; O08647;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98123114; PubMed=9452463;
RA Gao Z., Garbers D.L.;
RT "Species diversity in the structure of zonadhesin, a sperm-specific
RT membrane protein containing multiple cell adhesion molecule-like
RT domains.";
RL J. Biol. Chem. 273:3415-3421(1998).
RN [2]
RP SEQUENCE OF 4864-5376 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD.
CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -!- DOMAIN: THE VWFD DOMAIN 2 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -!- SIMILARITY: CONTAINS 3 MAM DOMAINS.
CC -!- SIMILARITY: CONTAINS 25 VWFD DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC ---

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U97068; AAC26680.1; -.
DR   EMBL; U83190; AAC53125.1; -.
DR   MGD; MGI:106656; Zan.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR000998; MAM.
DR   InterPro; IPR002919; TIL.
DR   InterPro; IPR003328; TILa.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00629; MAM; 3.
DR   Pfam; PF01826; TIL; 25.
DR   Pfam; PF02345; TILa; 25.
DR   Pfam; PF00094; vwd; 4.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00274; FOLN; 11.
DR   SMART; SM00137; MAM; 2.
DR   SMART; SM00214; VWC; 17.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 18.
DR   PROSITE; PS00740; MAM_1; FALSE_NEG.
DR   PROSITE; PS50060; MAM_2; 3.
KW   Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW   Repeat.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   5376       ZONADHESIN.
FT   DOMAIN       18   5310       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   5311   5337       POTENTIAL.
FT   DOMAIN     5338   5376       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       45    210       MAM 1.
FT   DOMAIN      215    374       MAM 2.
FT   DOMAIN      377    542       MAM 3.
FT   DOMAIN      547   1170       80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT                                (MUCIN-LIKE DOMAIN).
FT   DOMAIN     1171   1280       VWFD 1 (PARTIAL).
FT   DOMAIN     1281   1669       VWFD 2.
FT   DOMAIN     1670   2056       VWFD 3.
FT   DOMAIN     2057   2459       VWFD 4.
FT   DOMAIN     2460   2579       VWFD 5 (PARTIAL).
FT   DOMAIN     2580   2699       VWFD 6 (PARTIAL).
FT   DOMAIN     2700   2819       VWFD 7 (PARTIAL).
FT   DOMAIN     2820   2939       VWFD 8 (PARTIAL).
FT   DOMAIN     2940   3059       VWFD 9 (PARTIAL).
FT   DOMAIN     3060   3179       VWFD 10 (PARTIAL).
FT   DOMAIN     3180   3299       VWFD 11 (PARTIAL).
FT   DOMAIN     3300   3416       VWFD 12 (PARTIAL).
FT   DOMAIN     3417   3536       VWFD 13 (PARTIAL).
FT   DOMAIN     3537   3656       VWFD 14 (PARTIAL).
FT   DOMAIN     3657   3776       VWFD 15 (PARTIAL).
FT   DOMAIN     3777   3892       VWFD 16 (PARTIAL).
FT   DOMAIN     3893   4928       VWFD 17 (PARTIAL).
FT   DOMAIN     4029   4148       VWFD 18 (PARTIAL).
FT   DOMAIN     4149   4263       VWFD 19 (PARTIAL).
FT   DOMAIN     4264   4283       VWFD 20 (PARTIAL).
FT   DOMAIN     4384   4503       VWFD 21 (PARTIAL).
FT   DOMAIN     4504   4623       VWFD 22 (PARTIAL).
FT   DOMAIN     4624   4743       VWFD 23 (PARTIAL).
FT   DOMAIN     4744   4863       VWFD 24 (PARTIAL).
FT   DOMAIN     4864   5261       VWFD 25.
FT   DOMAIN     5259   5295       EGF-LIKE.
FT   DISULFID   5263   5274       BY SIMILARITY.
FT   DISULFID   5268   5283       BY SIMILARITY.
FT   DISULFID   5285   5294       BY SIMILARITY.
FT   CARBOHYD    339    339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1216   1216       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1239   1239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1314   1314       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1814   1814       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1908   1908       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1933   1933       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2028   2028       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2111   2111       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2142   2142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2332   2332       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2533   2533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2575   2575       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2692   2692       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2812   2812       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3052   3052       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3065   3065       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3144   3144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3172   3172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3288   3288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3292   3292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3782   3782       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4005   4005       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4136   4136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4243   4243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4254   4254       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4335   4335       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4376   4376       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4586   4586       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5136   5136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5252   5252       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   5376 AA;  579908 MW;  0E44DB77DF2A2620 CRC64;


  Query Match             8.3%;  Score 82.5;  DB 1;  Length 5376;
  Best Local Similarity   24.9%;  Pred. No. 33;
  Matches   43;  Conservative   12;  Mismatches   53;  Indels   65;  Gaps   10;

Qy      24 GTS-TASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSY 82
           |||  | | |       :||                  ||  :|  | |    : |
Db    4530 GTSPEAPSTC------EEGCV---------------CEPDYVLSNDKCVPSSEC----- 4562

Qy      83 HVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAW--RNACVNTSCGSASVSSYANTAGN 140
                 |:| ||    :|  |           |  |    | :|   :|  :
Db    4563 ------GCKDAHGV----LIPES---------KTWVSRGCTKNCTCKGGTVQCHD----- 4598

Qy     141 VYYRSPSNSYLYDNNLINTNCV-----LTKFSLLSGCSPSPAPDVSS----CG 184
             :  |: |   |||  |:|||         || : | ||  |  |  |     ||
Db    4599 --FSCPTGSRCLDNNEGNSNCVTYALKCPAHSLYTNCLPSCLPSCSDPEGLCG 4649


RESULT   6
PMP9_CHLPN
ID   PMP9_CHLPN     STANDARD;      PRT;   928 AA.
AC   Q9Z398;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable outer membrane protein pmp9 precursor (Polymorphic membrane
DE   protein 9) (Outer membrane protein 10).
GN   PMP9 OR OMP10 OR CPN0447 OR CP0306.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VR1310;
RX   MEDLINE=20007584; PubMed=10539856;
RA   Christiansen G., Boesen T., Hjerno K., Daugaard L., Mygind P.,
RA   Madsen A.S., Knudsen K., Falk E., Birkelund S.;
RT   "Molecular biology of Chlamydia pneumoniae surface proteins and their
RT   role in immunopathogenicity.";
RL   Am. Heart J. 138:S491-S495(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
```

```
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ133034; CAB37069.1; -.
DR   EMBL; AE001628; AAD18591.1; -.
DR   EMBL; AE002192; AAF38163.1; -.
DR   EMBL; AP002546; BAA98655.1; -.
DR   TIGR; CP0306; -.
DR   InterPro; IPR003368; DUF145.
DR   InterPro; IPR003357; OMP.
DR   Pfam; PF02415; DUF145; 1.
DR   Pfam; PF02385; OMP; 1.
KW   Outer membrane; Signal; Multigene family; Complete proteome.
FT   SIGNAL       1     26       POTENTIAL.
FT   CHAIN       27    928       PROBABLE OUTER MEMBRANE PROTEIN PMP9.
SQ   SEQUENCE   928 AA;  98332 MW;  58910A8F04F12219 CRC64;
```

```
  Query Match             8.3%;  Score 82;  DB 1;  Length 928;
  Best Local Similarity   21.8%;  Pred. No. 4.9;
  Matches   49;  Conservative   35;  Mismatches   69;  Indels   72;  Gaps   13;

Qy      10 TLNSLFNQIKNQSCGTSTA-SSPCITFRYPVDGCYARAHKMRQILMN--NGYDCE----- 61
           |: :|   :   : |: || :: |  |:        : |  : :| |   | :|
Db      57 TIYNLTGDVSITNAGSPTALTASC--FKETTGNLSFQGHGYQFLLQNIDAGANCTFTNTA 114

Qy      62 -------KQFVY------------GNLKASTGTCCVAWSYHVAILVSYKNASG------ 95
                  | |              | :| ||| | :  :|      :: | :|
Db     115 ANKLLSFSGFSYLSLIQTTNATTGTGAIK-STGACSIQSNYSCYFGQNFSNDNGGALQGS 173

Qy      96 ----------------VTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVS--SYANT 137
                           |:|     :|:|:| :|        :| :  ||| |  : ||
Db     174 SISLSLNPNLTFAKNKATQK---GGALYSTGGIT--------INNTLNSASFSENTAANN 222

Qy     138 AGNVYYRSPSNSYLYDNNLIN-TNCVLTKFSLLSG---CSPSPAP 178
           | :|  : ::|::  |  |:  |  :|  |   |   || : ||
Db     223 GGAIY--TEASSFISSNKAISFINNSVTATSATGGAIYCSSTSAP 265
```

```
RESULT   7
```

```
G156_PARPR
ID   G156_PARPR     STANDARD;      PRT;  2715 AA.
AC   P13837;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   156G surface protein precursor.
GN   156G.
OS   Paramecium primaurelia.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC   Paramecium.
OX   NCBI_TaxID=5886;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=156;
RX   MEDLINE=87060934; PubMed=3783679;
RA   Prat A., Katinka M., Caron F., Meyer E.;
RT   "Nucleotide sequence of the Paramecium primaurelia G surface protein.
RT   A huge protein with a highly periodic structure.";
RL   J. Mol. Biol. 189:47-60(1986).
CC   -!- FUNCTION: THIS PROTEIN IS THE SURFACE ANTIGEN OR IMMOBILIZATION
CC       ANTIGEN OF PARAMECIUM PRIMAURELIA.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- MISCELLANEOUS: IT HAS INTERNAL HOMOLOGIES AND A HIGHLY PERIODIC
CC       STRUCTURE WITH 37 PERIODS OF ABOUT 75 RESIDUES, EACH PERIOD
CC       CONTAINING 8 CYSTEINES, EXCEPT FOR FOUR HALF PERIODS. A VARIABLE
CC       PART OF 475 RESIDUES COMPRISES 4 ALMOST IDENTICAL PERIODS IN THE
CC       MIDDLE OF THE PROTEIN.
CC   -!- MISCELLANEOUS: EXPRESSION OF G PROTEIN OCCURS AT LOW TEMPERATURES
CC       (14-32 DEGREES CELSIUS).
CC   -!- SIMILARITY: 98% TO THE ALLELIC FORM 168G PROTEIN (AC P17053) IN
CC       PERIODIC STRUCTURE AND 80% IN VARIABLE DOMAIN IN THE MIDDLE OF
CC       THE PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03882; CAA27514.1; -.
DR   PIR; A23475; A23475.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR002895; Paramecium_SA.
DR   Pfam; PF01508; Paramecium_SA; 33.
KW   Signal; Repeat; Antigen; Membrane; GPI-anchor.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   2715       156G SURFACE PROTEIN.
FT   DOMAIN     106   2560       37 X 75 AA APPROXIMATE REPEATS.
FT   SIMILAR      1    222       88% TO PARAMECIUM TETRAURELIA A
FT                               PROTEIN.
SQ   SEQUENCE   2715 AA;  279551 MW;  97BE359AB9C7C298 CRC64;
```

```
  Query Match             8.3%;  Score 82;  DB 1;  Length 2715;
  Best Local Similarity   20.5%;  Pred. No. 17;
  Matches   51;  Conservative   23;  Mismatches   71;  Indels  104;  Gaps   11;

Qy      11 LNSLFNQIKNQSC----------------GTSTASSPCITFRYPVDG------------ 41
           :  ::|::|  ||                 : |    | |||   ||
Db     212 IQCVWNKLKTTSCYWDGAACKDRICDNAPTSLTTDDACKTFR--TDGTCTTKANGGCVTR 269

Qy      42 --CYARAHKMRQILMNNGYDC--------EKQ----------------FVYGNLKASTGT 75
             | |   :   |  ::| ||        :|                || | :  | |
Db     270 TTCAAATIQASCIKNSSGGDCYWTGTACVDKACANTPTTIATNSACAGFVTGCITKSGGG 329

Qy      76 CCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASV---- 131
           | |  :  ||          |    : :|| |      |  |   |   :| :||
Db     330 CVVNGACSVA---------NVQAACVKNPSNF------DCIWDTTCKEKTCANASTTNNT 374

Qy     132 ----SSYANT-----AGNVYYRSPSN------------SYLYDNNLIN-------TNCVL 163
```

```
           :|| :|      |    |: :|           :|   || |        ||
Db    375 HDLCTSYLSTCTVKSGGGCQNRTCANAPTTMTTNDACEAYFTGNNCITKSGGGCVTNTTC 434

Qy    164 TKFSLLSGC 172
          :| : |
Db    435 AAITLEAAC 443


RESULT   8
GPI_ZINOF
ID   GPI_ZINOF      STANDARD;      PRT;   221 AA.
AC   P82473;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Cysteine proteinase GP-I (EC 3.4.22.-).
OS   Zingiber officinale (Ginger).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;
OC   Zingiber.
OX   NCBI_TaxID=94328;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Root;
RX   MEDLINE=20156257; PubMed=10691991;
RA   Choi K.H., Laursen R.A.;
RT   "Amino-acid sequence and glycan structures of cysteine proteases with
RT   proline specificity from ginger rhizome Zingiber officinale.";
RL   Eur. J. Biochem. 267:1516-1526(2000).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE WITH A PROLINE RESIDUE
CC       AT P2.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC       PAPAIN FAMILY OF THIOL PROTEASES.
CC   -!- CAUTION: THE AUTHORS REGARD THE SEQUENCE AS TENTATIVE, AS THEY
CC       BELIEVE THAT IT MAY HAVE BEEN CONTAMINATED BY A HOMOLOGOUS
CC       PROTEIN.
DR   HSSP; P00785; 2ACT.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Thiolprot_act_site.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; FALSE_NEG.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW   Hydrolase; Thiol protease; Glycoprotein.
FT   ACT_SITE     27     27       BY SIMILARITY.
FT   ACT_SITE    161    161       BY SIMILARITY.
FT   DISULFID     24     65       BY SIMILARITY.
FT   DISULFID     58     98       BY SIMILARITY.
FT   DISULFID    155    206       BY SIMILARITY.
FT   CARBOHYD     95     95       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    156    156       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   UNSURE        2      2       OR D.
SQ   SEQUENCE   221 AA;  24241 MW;  3035D7870EA743DB CRC64;


  Query Match             8.0%;  Score 79;  DB 1;  Length 221;
  Best Local Similarity   22.5%;  Pred. No. 1.8;
  Matches   36;  Conservative   19;  Mismatches   57;  Indels   48;  Gaps    8;

Qy     18 IKNQ-SCGTSTASSPCITFRYPVDG------------------CYARAH---------K 48
          :|||  ||:  |          |:|                    |  | |
Db     18 VKNQGGCGSCWAFDAIAA----VEGINQIVTGDLISLSEQQLVDCSTRNHGCEGGWPYRA 73

Qy     49 MRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFS 108
           : |: | | : |: : |     : |||    : ||  : ||:|     || :      :
Db     74 FQYIINNGGINSEEHYPY---TGTNGTCDTKENAHVVSIDSYRNVPSNDEKSL--QKAVA 128

Qy    109 SGPVTDT---------AWRNACVNTSCGSASVSSYANTAG 139
          : ||: |          :||     || : | : |    |
Db    129 NQPVSVTMDAAGRDFQLYRNGIFTGSC-NISANHYRTVGG 167
```

```
RESULT   9
SPM1_SCHPO
ID   SPM1_SCHPO     STANDARD;      PRT;   422 AA.
AC   Q92398;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mitogen-activated protein kinase spm1 (EC 2.7.1.-) (MAP kinase spm1)
DE   (MAP kinase pmk1).
GN   SPM1 OR PMK1 OR SPBC119.08.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97098653; PubMed=8943330;
RA   Toda T., Dhut S., Superti-Furga G., Gotoh G., Nishida E., Sugiura R.,
RA   Kuno T.;
RT   "The fission yeast pmk1+ gene encodes a novel mitogen-activated
RT   protein kinase homolog which regulates cell integrity and functions
RT   coordinately with the protein kinase C pathway.";
RL   Mol. Cell. Biol. 16:6752-6764(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97280820; PubMed=9135147;
RA   Zaitsevskaya-Carter T., Cooper J.A.;
RT   "Spm1, a stress-activated MAP kinase that regulates morphogenesis in
RT   S.pombe.";
RL   EMBO J. 16:1318-1331(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REGULATES CELL INTEGRITY AND FUNCTIONS COORDINATELY WITH
CC       THE PROTEIN KINASE C PATHWAY (PCK1 AND PCK2). INVOLVED THE
CC       REGULATION OF WALL ARCHITECTURE, CELL SHAPE, CYTOKINESIS IN
CC       EXPONENTIAL AND STATIONARY PHASE, AND METABOLISM OF IONS.
CC   -!- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE
CC       PHOSPHORYLATION BY SKH1/PEK1.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       MAP KINASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X98243; CAA66899.1; -.
DR   EMBL; U65405; AAC49707.1; -.
DR   EMBL; AL022117; CAA17923.1; -.
DR   HSSP; Q16539; 1WFC.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003527; MAP_kin.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS01351; MAPK; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Cell cycle; Phosphorylation.
FT   DOMAIN       21    314       PROTEIN KINASE.
FT   NP_BIND      27     35       ATP (BY SIMILARITY).
FT   BINDING      52     52       ATP (BY SIMILARITY).
FT   ACT_SITE    149    149       BY SIMILARITY.
FT   MOD_RES     186    186       PHOSPHORYLATION (ACTIVATES THE KINASE)
```

```
FT                                (BY SIMILARITY).
FT   MOD_RES      188     188     PHOSPHORYLATION (ACTIVATES THE KINASE)
FT                                (BY SIMILARITY).
FT   DOMAIN       389     394     POLY-SER.
SQ   SEQUENCE   422 AA;  48261 MW;  FD02521E64E8BF82 CRC64;


  Query Match             7.8%;  Score 77.5;  DB 1;  Length 422;
  Best Local Similarity   24.4%;  Pred. No. 5.2;
  Matches   29;  Conservative   17;  Mismatches   36;  Indels   37;  Gaps    6;

Qy      84 VAILVSYKNASGVT---EKRIIDPSLF------------------SSGPVTDTAWRNAC 121
           : :|: ::|   :|   :  ||:|  |                  |  |:||  :::
Db      71 IKLLIHFRNHRNITCIYDLDIINPYNFNEVYIYEELMEADLNAIIKSGQPLTDAHFQSFI 130

Qy     122 VNTSCGSASVSSYANTAGNVYYR--SPSNSYLYDNNLINTNCVL--TKFSLLSGCSPSP 176
             ||   : |      || :|   | |       |:| :| |    | |  ||| :|
Db     131 YQILCGLKYIHS-----ANVIHRDLKPGNL------LVNADCELKICDFGLARGCSENP 178


RESULT  10
PKP1_MOUSE
ID   PKP1_MOUSE     STANDARD;      PRT;   728 AA.
AC   P97350;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Plakophilin 1.
GN   PKP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Skin;
RA   Nimmrich V., Hunziker A.H., Franke W.W.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y07941; CAA69240.1; -.
DR   MGD; MGI:1328359; Pkp1.
DR   InterPro; IPR000225; Armadillo.
DR   Pfam; PF00514; Armadillo_seg; 5.
DR   SMART; SM00185; ARM; 3.
DR   PROSITE; PS50176; ARM_REPEAT; 3.
KW   Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
KW   Repeat.
FT   REPEAT       235     275     ARM 1.
FT   REPEAT       278     317     ARM 2.
FT   REPEAT       320     360     ARM 3.
FT   REPEAT       419     464     ARM 4.
FT   REPEAT       517     557     ARM 5.
FT   REPEAT       565     604     ARM 6.
FT   REPEAT       606     650     ARM 7.
SQ   SEQUENCE   728 AA;  80896 MW;  BDAC5BA7B4118AC0 CRC64;


  Query Match             7.8%;  Score 77.5;  DB 1;  Length 728;
  Best Local Similarity   26.1%;  Pred. No. 9.8;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   25;  Gaps    6;

Qy      41 GCYA-RAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEK 99
           ||:: |  ||    ||| |||       | | |      :| ||   : :| |  | ::|
Db     492 GCFSNRGDKM----MNNNYDCPLPEEETNPKGS------SWLYHSDAIRTYLNLMGKSKK 541

Qy     100 RIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRS-PSNSYLYDNNLIN 158
                   |  |   |  | :     :|:  :    :  :  |  : |  :   |
Db     542 ----------DATLEACAGALQNLTASKGLMSNGMSQLIGIKEKGLPQIARLLQSG--N 588

Qy     159 TNCVLTKFSLLSGCSPSP 176
           :: | :  ||||  |  |
Db     589 SDVVRSGASLLSNMSRHP 606


RESULT  11
ABFB_ASPNG
ID   ABFB_ASPNG     STANDARD;      PRT;   499 AA.
AC   P42255;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Alpha-L-arabinofuranosidase B precursor (EC 3.2.1.55) (Arabinosidase
DE   B) (ABF B).
GN   ABFB.
OS   Aspergillus niger.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5061;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 19-30; 221-234; 285-304 AND
RP   312-324.
RC   STRAIN=CBS 120.49 / N400;
RX   MEDLINE=94130338; PubMed=8299175;
RA   Flipphi M.J.A., van Heuvel M., van der Veen P., Visser J.,
RA   de Graaff L.H.;
RT   "Cloning and characterization of the abfB gene coding for the major
RT   alpha-L-arabinofuranosidase (ABF B) of Aspergillus niger.";
RL   Curr. Genet. 24:525-532(1993).
CC   -!- FUNCTION: ABLE TO HYDROLYSE 1,5-, 1,3- AND 1,2-ALPHA-LINKAGES NOT
CC       ONLY IN L-ARABINOFURANOSYL OLIGOSACCHARIDES, BUT ALSO IN POLYSAC-
CC       CHARIDES CONTAINING TERMINAL NON-REDUCING L-ARABINOFURANOSES IN
CC       SIDE CHAINS, LIKE L-ARABINAN, ARABINOGALACTAN AND ARABINOXYLAN.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
CC       arabinofuranoside residues in alpha-L-arabinosides.
CC   -!- PATHWAY: INVOLVED IN DEGRADATION OF THE PLANT CELL WALL
CC       POLYSACCHARIDE L-ARABINAN.
CC   -!- INDUCTION: BY GROWTH ON POLYMERIC SUBSTRATES AND L-ARABITOL.
CC   -!- SIMILARITY: BELONGS TO FAMILY 54 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74777; CAA52785.1; -.
DR   EMBL; L23502; AAB53944.1; -.
DR   EMBL; A27977; CAA01903.1; -.
KW   Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1      18
FT   CHAIN        19     499       ALPHA-L-ARABINOFURANOSIDASE B.
FT   CARBOHYD     83      83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    202     202       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   499 AA;  52523 MW;  F1F6C3B47AA419C9 CRC64;


  Query Match             7.7%;  Score 76.5;  DB 1;  Length 499;
  Best Local Similarity   24.7%;  Pred. No. 7.9;
  Matches   59;  Conservative   22;  Mismatches   77;  Indels   81;  Gaps   14;
```

```
Qy       1 LASVIPDVATLN-----SLF-----NQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMR 50
           ||| |    |||     :|         |:: ||:|   |        :| ||
Db     120 LASAIGAPVTLNGQKAYGVFMSPGTGYRNNEATGTATGDEP--------EGMYA------ 165

Qy      51 QILMNNGYDCEKQFVYGNLKAS---TGT---------CCVAWSYHVA----ILVSYKN-- 92
            :|    |:    | ||| : |   ||              |:|       |:|  :|
Db     166 -VLDGTHYNDACCFDYGNAETSSTDTGAGHMEAIYLGNSTTWGYGAGDGPWIMVDMENNL 224

Qy      93 ASGVTE-KRIIDPSL---FSSGPVTDTAWRNACVNTSCGSASVSSY-------------- 134
           ||  |     |||:   | :  |   | : |     :  | |:|:|
Db     225 FSGADEGYNSGDPSISYSFVTAAVKGGADKWAIRGGNAASGSLSTYYSGARPDYSGYNPM 284

Qy     135 --------------ANTAGNVYYRSPSNS-YLYDN--NLINTNCVLTKF---SLLSGCS 173
                         :| |   :|     | |  |:  | :  | |  |:   ||:|| |
Db     285 SKEGAIILGIGGDNSNGAQGTFYEGVMTSGYPSDDVENSVQENIVAAKYVSGSLVSGPS 343


RESULT  12
IRK4_CAEEL
ID   IRK4_CAEEL     STANDARD;      PRT;   505 AA.
AC   P52192;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Inward rectifier potassium channel irk-4.
GN   IRK-4 OR R03E9.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wilcox L.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INWARD RECTIFIER K+ CHANNELS ARE CHARACTERIZED BY A
CC       GREATER TENDANCY TO ALLOW POTASSIUM TO FLOW INTO THE CELL RATHER
CC       THAN OUT OF IT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE INWARD RECTIFIER-TYPE K+ CHANNEL
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U40947; AAC48070.1; -.
DR   WormPep; R03E9.4; CE04787.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR001838; KIR_channel.
DR   Pfam; PF01007; IRK; 1.
DR   PRINTS; PR01320; KIRCHANNEL.
KW   Ionic channel; Ion transport; Voltage-gated channel; Transmembrane;
KW   Potassium transport.
FT   DOMAIN        1    114       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    115    135       POTENTIAL.
FT   DOMAIN      136    190       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    191    211       POTENTIAL.
FT   DOMAIN      212    505       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   505 AA;  57419 MW;  C1DF6FDB2B280434 CRC64;


  Query Match             7.7%;  Score 76.5;  DB 1;  Length 505;
  Best Local Similarity   22.1%;  Pred. No. 8;
  Matches   46;  Conservative   27;  Mismatches   78;  Indels   57;  Gaps    8;

Qy       6 PDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYD------ 59
           | :|  : ||       |    : ||   :              :| ||:  ::
Db     285 PSIADDDRLFLVWPTTLCHVIDSRSPLYNYN------------QQTLMSAQFEIIVLLE 331

Qy      60 --CEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLF-SSGPVTDTA 116
             |   :    | |     | | :    ||:|: ::|  :   ||  || |: ||   |
Db     332 GIVESTGMTAQAKTSYLPSEVLWGHRFRKLVTYQRSNGSYQ---IDYDLFHSTYPVRTPA 388

Qy     117 W--------------------------RNACVNTSCGSASVSSYANTAGNVYYRSPSNS 149
                                        :  :|   | |  || :|  | :::| |||
Db     389 MSPAEFYSSKPNLKDYYCHDSHEHKLEDNRSSDSTPLPSPSPYSYPSTPLN-HFQSSSNS 447

Qy     150 YLYDNN--LINTNCVLTKFSLLSGCSPS 175
             :: ||    ||  |  :  :|  | |:
Db     448 PVFSNNHSKFNTEAVTCEAGML--CPPT 473


RESULT  13
PKP1_HUMAN
ID   PKP1_HUMAN     STANDARD;      PRT;   747 AA.
AC   Q13835; Q15152; O00645;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Plakophilin 1 (Band-6-protein) (B6P).
GN   PKP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Epidermis;
RX   MEDLINE=95074299; PubMed=7527055;
RA   Hatzfeld M., Kristjansson G.I., Plessmann U., Weber K.;
RT   "Band 6 protein, a major constituent of desmosomes from stratified
RT   epithelia, is a novel member of the armadillo multigene family.";
RL   J. Cell Sci. 107:2259-2270(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Bosch A.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   PubMed=9369526;
RA   Schmidt A., Langbein L., Rode M., Praetzel S., Zimbelmann R.,
RA   Franke W.W.;
RT   "Plakophilins 1a and 1b: widespread nuclear proteins recruited in
RT   specific epithelial cells as desmosomal plaque components.";
RL   Cell Tissue Res. 290:481-499(1997).
CC   -!- FUNCTION: SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Isoform 1 is also associated with
CC       desmosomes.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/A AND 2/B (SHOWN HERE); ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: NUCLEAR ISOFORM IS WIDELY EXPRESSED. ISOFORM 1
CC       IS EXPRESSED IN STRATIFIED SQUAMOUS, COMPLEX, GLANDULAR DUCT AND
CC       BLADDER EPITHELIA.
CC   -!- DISEASE: DEFECTS IN PKP1 ARE ASSOCIATED WITH ECTODERMAL
CC       DYSPLASIA/SKIN FRAGILITY SYNDROME.
CC   -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 8 ARM REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79293; CAA55881.1; -.
DR   EMBL; Z34974; CAA84426.1; -.
DR   EMBL; Z73677; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z73678; CAA98022.1; -.
DR   MIM; 601975; -.
```

```
DR   MIM; 604536; -.
DR   InterPro; IPR000225; Armadillo.
DR   Pfam; PF00514; Armadillo_seg; 5.
DR   SMART; SM00185; ARM; 3.
DR   PROSITE; PS50176; ARM_REPEAT; 3.
KW   Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
KW   Repeat; Alternative splicing.
FT   REPEAT      234    274       ARM 1.
FT   REPEAT      277    316       ARM 2.
FT   REPEAT      319    359       ARM 3.
FT   REPEAT      439    484       ARM 4.
FT   REPEAT      537    577       ARM 5.
FT   REPEAT      585    624       ARM 6.
FT   REPEAT      626    670       ARM 7.
FT   REPEAT      673    713       ARM 8.
FT   VARSPLIC    412    432       MISSING (IN ISOFORM 1).
FT   CONFLICT    154    154       R -> G (IN REF. 1).
FT   CONFLICT    216    222       PPISCNK -> RHLLQQ (IN REF. 1).
FT   CONFLICT    462    462       V -> E (IN REF. 1).
FT   CONFLICT    496    496       Q -> K (IN REF. 1).
FT   CONFLICT    506    506       T -> P (IN REF. 1).
FT   CONFLICT    553    553       L -> S (IN REF. 1).
SQ   SEQUENCE   747 AA;  82860 MW;  60C1BCCC50AB4E5F CRC64;


  Query Match             7.7%;  Score 76.5;  DB 1;  Length 747;
  Best Local Similarity   26.8%;  Pred. No. 13;
  Matches   37;  Conservative   16;  Mismatches   60;  Indels   25;  Gaps    6;

Qy      41 GCYA-RAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEK 99
           ||:: :: ||    ||| |||       | | |       | ||   : :| |  | ::|
Db     512 GCFSNKSDKM----MNNNYDCPLPEEETNPKGS------GWLYHSDAIRTYLNLMGKSKK 561

Qy     100 RIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSS-YANTAGNVYYRSPSNSYLYDNNLIN 158
                   | |   |  | :    :|| :   |      |  : |  :   |
Db     562 -----------DATLEACAGALQNLTASKGLMSSGMSQLIGLKEKGLPQIARLLQSG--N 608

Qy     159 TNCVLTKFSLLSGCSPSP 176
           :: | :  ||||  |  |
Db     609 SDVVRSGASLLSNMSRHP 626


RESULT  14
G168_PARPR
ID   G168_PARPR     STANDARD;      PRT;  2704 AA.
AC   P17053;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   168G surface protein precursor.
GN   168G.
OS   Paramecium primaurelia.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC   Paramecium.
OX   NCBI_TaxID=5886;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=90172419; PubMed=2308165;
RA   Prat A.;
RT   "Conserved sequences flank variable tandem repeats in two alleles of
RT   the G surface protein of Paramecium primaurelia.";
RL   J. Mol. Biol. 211:521-535(1990).
CC   -!- FUNCTION: THIS PROTEIN IS THE SURFACE ANTIGEN OR IMMOBILIZATION
CC       ANTIGEN OF PARAMECIUM PRIMAURELIA.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- MISCELLANEOUS: IT HAS INTERNAL HOMOLOGIES AND A HIGHLY PERIODIC
CC       STRUCTURE WITH 37 PERIODS OF ABOUT 75 RESIDUES, EACH PERIOD
CC       CONTAINING 8 CYSTEINES, EXCEPT FOR FOUR HALF PERIODS. A VARIABLE
CC       PART OF 475 RESIDUES COMPRISES 4 ALMOST IDENTICAL PERIODS IN THE
CC       MIDDLE OF THE PROTEIN.
CC   -!- MISCELLANEOUS: EXPRESSION OF G PROTEIN OCCURS AT LOW TEMPERATURES
CC       (14-32 DEGREES CELSIUS).
CC   -!- SIMILARITY: 98% TO THE ALLELIC FORM 156G PROTEIN (AC P13837) IN
CC       PERIODIC STRUCTURE AND 80% IN THE VARIABLE DOMAIN IN THE MIDDLE
CC       OF THE PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52133; CAA36378.1; -.
DR   PIR; S09118; S09118.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR002895; Paramecium_SA.
DR   Pfam; PF01508; Paramecium_SA; 33.
KW   Signal; Repeat; Antigen; Membrane; GPI-anchor.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   2704       168G SURFACE PROTEIN.
FT   DOMAIN      106   2560       37 X 75 AA APPROXIMATE REPEATS.
FT   DOMAIN     1060   1424       VARIABLE DOMAIN, COMPRISES 4 ALMOST
FT                                IDENTICAL REPEATS.
SQ   SEQUENCE   2704 AA;  278775 MW;  40EA0A0B18EE2119 CRC64;


  Query Match             7.7%;  Score 76;  DB 1;  Length 2704;
  Best Local Similarity   19.9%;  Pred. No. 61;
  Matches   50;  Conservative   22;  Mismatches   71;  Indels  108;  Gaps   11;

Qy      11 LNSLFNQIKNQSC----------------GTSTASSPCITFRYPVDG------------ 41
           :  ::|::|  ||                 : |    | |||   ||
Db     206 IQCVWNKLKTTSCYWDGAACKDRICDNAPTSLTTDDACKTFR--TDGTCTTKANGGCVTR 263

Qy      42 --CYARAHKMRQILMNNGYDC--------EKQ----------------FVYGNLKASTGT 75
             | |   :   |  ::| ||        :|                  || | :  | |
Db     264 TTCAAATIQASCIKNSSGGDCYWTGTACVDKTCANAPTTMTTNSACAGFVTGCITKSGGG 323

Qy      76 CCVAWSYHVAIL--VSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASV-- 131
           |    :  || :     ||:|                  |  |   |   :| :|
Db     324 CVANGACSVANVQAACVKNSSNF----------------DCIWDTTCKEKTCANAPTTN 366

Qy     132 ------SSYANT-----AGNVYYRSPSN-----------SYLYDNNLIN-------TNC 161
                :|| :|      |    || :|            :||  || |        ||
Db     367 NTHDLCTSYLSTCTVKSGGGCQNRSCANAPTTMTTNDACEAYLTGNNCITKSGGGCVTNT 426

Qy     162 VLTKFSLLSGC 172
           :| : |
Db     427 TCAAITLEAAC 437


RESULT  15
CFAI_MOUSE
ID   CFAI_MOUSE     STANDARD;      PRT;   603 AA.
AC   Q61129; Q9WU07;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor I precursor (EC 3.4.21.45) (C3B/C4B inactivator).
GN   IF OR CFI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=96175003; PubMed=8604219;
RA   Minta J.O., Wong M.J., Kozak C.A., Kunnath-Muglia L.M., Goldberger G.;
RT   "cDNA cloning, sequencing and chromosomal assignment of the gene for
RT   mouse complement factor I (C3b/C4b inactivator): identification of a
RT   species specific divergent segment in factor I.";
RL   Mol. Immunol. 33:101-112(1996).
```

```
RN   [2]
RP   SEQUENCE OF 114-334 FROM N.A.
RC   STRAIN=129/SV; TISSUE=Kidney;
RX   MEDLINE=99220287; PubMed=10204086;
RA   Yun Y.-S., Goldberger G., Minta J.O.;
RT   "Cloning and characterization of the non-catalytic heavy chain of
RT   mouse complement factor I gene: structure comparison with the human
RT   homologue.";
RL   Biochem. Mol. Biol. Int. 47:493-500(1999).
CC   -!- FUNCTION: RESPONSIBLE FOR CLEAVING THE ALPHA-CHAINS OF C4B AND C3B
CC       IN THE PRESENCE OF THE COFACTORS C4-BINDING PROTEIN AND FACTOR H
CC       RESPECTIVELY.
CC   -!- SUBUNIT: HETERODIMER OF A LIGHT AND HEAVY CHAINS LINKED BY
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U47810; AAB00438.1; -.
DR   EMBL; AF100565; AAD32965.1; -.
DR   EMBL; AF100559; AAD32965.1; JOINED.
DR   EMBL; AF100560; AAD32965.1; JOINED.
DR   EMBL; AF100561; AAD32965.1; JOINED.
DR   EMBL; AF100562; AAD32965.1; JOINED.
DR   EMBL; AF100563; AAD32965.1; JOINED.
DR   EMBL; AF100564; AAD32965.1; JOINED.
DR   HSSP; P00750; 1RTF.
DR   MEROPS; S01.199; -.
DR   MGD; MGI:105937; Cfi.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR003884; FacI_MAC.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR001190; SRCR.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR002350; kazal.
DR   Pfam; PF00050; kazal; 1.
DR   Pfam; PF00057; ldl_recept_a; 1.
DR   Pfam; PF00530; SRCR; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00057; FIMAC; 1.
DR   SMART; SM00280; KAZAL; 1.
DR   SMART; SM00192; LDLa; 2.
DR   SMART; SM00202; SR; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS01209; LDLRA_1; 1.
DR   PROSITE; PS50068; LDLRA_2; 2.
DR   PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR   PROSITE; PS50287; SRCR_2; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease;
KW   Repeat; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    603       COMPLEMENT FACTOR I.
FT   CHAIN        19    356       COMPLEMENT FACTOR I HEAVY CHAIN.
FT   CHAIN       361    603       COMPLEMENT FACTOR I LIGHT CHAIN.
FT   DOMAIN       63    109       KAZAL-LIKE.
FT   DOMAIN      117    217       SRCR.
FT   DOMAIN      218    262       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      263    299       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      361    603       SERINE PROTEASE.
FT   DISULFID    234    252       BY SIMILARITY.
FT   DISULFID    246      ?       BY SIMILARITY.
FT   DISULFID    264    276       BY SIMILARITY.
FT   DISULFID    271    289       BY SIMILARITY.
FT   DISULFID    283    298       BY SIMILARITY.
FT   DISULFID    386    402       BY SIMILARITY.
FT   DISULFID    487    551       BY SIMILARITY.
FT   DISULFID    515    530       BY SIMILARITY.
FT   DISULFID    541    570       BY SIMILARITY.
FT   CARBOHYD    106    106       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    267    267       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    514    514       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   ACT_SITE    401    401       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    449    449       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    545    545       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CONFLICT    114    114       N -> K (IN REF. 2).
FT   CONFLICT    236    252       NGKHIPQEKACNGVNDC -> MGSTFLRRKPATVSMTV
FT                                (IN REF. 1).
SQ   SEQUENCE   603 AA;  67234 MW;  E2C1D43261C719E8 CRC64;


  Query Match             7.6%;  Score 75.5;  DB 1;  Length 603;
  Best Local Similarity   18.5%;  Pred. No. 12;
  Matches   35;  Conservative   35;  Mismatches   70;  Indels   49;  Gaps   10;

Qy       2 ASVIP--DVATLNSLFNQIKNQSC------------GTSTASSP--CITFRYPVDGCYA 44
           || :|  ::     |  :  ::||             ||     |  |     ||
Db      23 ASDLPQEELVDQKCLLQKYTHRSCNKVFCQPWQRCIEGTCICKLPYQCPRAGTPVCAMNG 82

Qy      45 RAHKMRQILMNNGYDCEKQF--VYGNLKAS-TGTCCVAWSYHVAILVSYKNASGVTEKRI 101
           |::          |  :| |  ::  :| |  |||    :::|:::       |: : ::
Db      83 RSYPT--------YCHQKSFECLHPEIKFSHNGTCAAEGNFNVSLIYGRTKTEGLVQVKL 134

Qy     102 IDPS--LFSSGPVTDTAWRNACVNTSC-------GSASVSSYANTAGNVYYRSPSNSYLY 152
           :|    :|    :   :|  |  | :|       |   :    | :||:        ::
Db     135 VDQDERMF----ICKNSWSMAEANVACVDLGFPLGVRDIQGSFNISGNL--------HIN 182

Qy     153 DNNLINTNC 161
           |   :: :|
Db     183 DTECLHVHC 191


Search completed: June 27, 2002, 21:59:19
Job time: 662 sec
```

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:47:27 ; Search time 102.74 Seconds
(without alignments)
311.505 Million cell updates/sec

Title: US-09-727-769A-6
Perfect score: 991
Sequence: 1 LASVIPDVATLNSLFNQIKN.....FSLLSGCSPSPAPDVSSCGF 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	991	100.0	320	2	Q9AQQ8	Q9aqq8 chryseobact
2	96.5	9.7	534	17	O26459	O26459 methanother
3	92	9.3	551	16	Q9HYT4	Q9hyt4 pseudomonas
4	92	9.3	2395	5	Q27167	Q27167 paramecium
5	91	9.2	594	10	Q9S746	Q9s746 arabidopsis
6	90	9.1	243	5	Q18804	Q18804 caenorhabdi
7	89.5	9.0	927	5	O02364	O02364 caenorhabdi
8	89	9.0	672	5	Q95QP7	Q95qp7 caenorhabdi
9	88.5	8.9	283	17	O26933	O26933 methanother
10	85.5	8.6	534	3	O94317	O94317 schizosacch
11	85	8.6	594	10	Q9SXZ3	Q9sxz3 arabidopsis
12	84	8.5	342	13	P70004	P70004 xenopus lae
13	84	8.5	3726	5	Q9VFL1	Q9vfl1 drosophila
14	83.5	8.4	619	5	Q9GTE0	Q9gte0 caenorhabdi
15	83.5	8.4	624	5	Q9GTD9	Q9gtd9 caenorhabdi
16	83.5	8.4	676	5	O45184	O45184 caenorhabdi
17	83	8.4	95	5	Q9N992	Q9n992 leishmania
18	83	8.4	930	5	O77151	O77151 cryptospori
19	82.5	8.3	5374	11	Q99ND0	Q99nd0 mus musculu
20	82	8.3	703	3	Q08844	Q08844 saccharomyc
21	81.5	8.2	361	5	Q9XUH3	Q9xuh3 caenorhabdi
22	81.5	8.2	1360	12	Q9IKD1	Q9ikd1 rat sialoda
23	81	8.2	346	5	Q9VW94	Q9vw94 drosophila
24	81	8.2	485	10	Q9LXL1	Q9lxl1 arabidopsis
25	80.5	8.1	508	10	Q9LTM8	Q9ltm8 arabidopsis
26	79.5	8.0	207	5	Q20608	Q20608 caenorhabdi
27	79.5	8.0	513	2	O24719	O24719 comamonas t
28	79.5	8.0	723	4	O95789	O95789 homo sapien
29	79.5	8.0	1316	17	Q9HLF6	Q9hlf6 thermoplasm
30	79.5	8.0	1894	11	P70206	P70206 mus musculu
31	79.5	8.0	1951	5	Q17042	Q17042 aequipecten
32	79	8.0	275	3	P78853	P78853 schizosacch
33	79	8.0	542	3	Q9P378	Q9p378 schizosacch
34	79	8.0	732	16	O31447	O31447 bacillus su
35	79	8.0	1290	5	Q9VTR8	Q9vtr8 drosophila
36	79	8.0	1314	5	Q9BLV9	Q9blv9 leishmania
37	78.5	7.9	370	3	O94737	O94737 pneumocysti
38	78.5	7.9	719	16	Q9CNW9	Q9cnw9 pasteurella
39	78.5	7.9	725	4	Q13835	Q13835 homo sapien
40	78.5	7.9	1067	5	O16249	O16249 caenorhabdi
41	77.5	7.8	400	5	Q95QP8	Q95qp8 caenorhabdi
42	77.5	7.8	539	12	Q91I41	Q91i41 avian infec
43	77.5	7.8	728	11	P97350	P97350 mus musculu
44	77.5	7.8	848	15	O40366	O40366 human immun
45	77.5	7.8	2717	5	Q94710	Q94710 paramecium

ALIGNMENTS

RESULT 1
Q9AQQ8
ID Q9AQQ8 PRELIMINARY; PRT; 320 AA.
AC Q9AQQ8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN-GLUTAMINASE PRECURSOR.
GN PRGA.
OS Chryseobacterium proteolyticum.
OC Bacteria; CFB group; Flavobacteria; Flavobacteriaceae;
OC Chryseobacterium.
OX NCBI_TaxID=118127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9670;
RX MEDLINE=21153247; PubMed=11231294;
RA Yamaguchi S., Jeenes D.J., Archer D.B.;
RT "Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme
RT that deamidates glutaminyl residues in proteins: purification,
RT characterization and gene cloning.";
RL Eur. J. Biochem. 268:1410-1421(2001).
DR EMBL; AB046594; BAB21508.1; -.
KW Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 136 320
SQ SEQUENCE 320 AA; 35044 MW; C67823D2BC131410 CRC64;

Query Match 100.0%; Score 991; DB 2; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.8e-90;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 LASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 LASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDC 195

Qy      61 EKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    196 EKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNA 255

Qy    121 CVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 CVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDV 315

Qy    181 SSCGF 185
          |||||
Db    316 SSCGF 320


RESULT   2
O26459
ID   O26459      PRELIMINARY;      PRT;   534 AA.
AC   O26459;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CONSERVED PROTEIN.
GN   MTH359.
OS   Methanothermobacter thermautotrophicus.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000821; AAB84865.1; -.
DR   InterPro; IPR002931; Transglut_core.
DR   Pfam; PF01841; Transglut_core; 1.
DR   SMART; SM00460; TGc; 1.
KW   Complete proteome.
SQ   SEQUENCE   534 AA;  57628 MW;  588412BD8D342E26 CRC64;


  Query Match             9.7%;  Score 96.5;  DB 17;  Length 534;
  Best Local Similarity   24.1%;  Pred. No. 0.24;
  Matches   39;  Conservative   23;  Mismatches   71;  Indels   29;  Gaps    4;

Qy      5 IPDVATLNSLFNQIKNQS-----CGTSTASSPCITFRYPVD----------------GCY 43
          :|: J:  | : :||:       | |||     :   |||                 :
Db    358 LPNTVTVTRLASSLKNRPENDPYRGESTARYLASSASCPVDSPDIRSLASEITRGLTSTF 417

Qy     44 ARAHK----MRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEK 99
          :||      :|  :  : |   |    | ||  || |      |  :||:   |:|:  :
Db    418 SRAEAVFGWVRDNINYSFYYNTKYGAVGTLKNRTGNCV----DHTHLLVALARAAGIPAR 473

Qy    100 RIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNV 141
          :     |:|| |    |    |  :  :|  :|  |: | |
Db    474 YVHGTCNFTSGNVYGHVWAQLLVGDTWYAADATSSRNSLGVV 515


RESULT   3
Q9HYT4
ID   Q9HYT4      PRELIMINARY;      PRT;   551 AA.
AC   Q9HYT4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HYPOTHETICAL PROTEIN PA3310.
GN   PA3310.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004753; AAG06698.1; -.
DR   InterPro; IPR003371; DUF146.
DR   Pfam; PF02418; DUF146; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   551 AA;  61421 MW;  3E3EAF5C308BE815 CRC64;


  Query Match             9.3%;  Score 92;  DB 16;  Length 551;
  Best Local Similarity   21.6%;  Pred. No. 0.71;
  Matches   42;  Conservative   24;  Mismatches   54;  Indels   74;  Gaps    9;

Qy     15 FNQIKNQSCGTSTA-SSPCITFRYPVDGC---YARAHK----------MRQILMNNGYDC 60
          | |::  ||||||| | ||:  :|| :      |: |:          :: : : |  ||
Db    274 FTQVR--SCGTSTAVSVPCMFSQYPREDYSDKKAKTHEGLLDILQRAGVQVLWLENNSDC 331

Qy     61 EKQFVYGNLKASTGTC--------------------CVAWSYHVAILVSYKNASGVTEK 99
          :            |||                    |:  |  | :    :   |: :
Db    332 K-----------GTCLRVPNRDIPKTQPSPFCDGKNCLDESLLVGL---QEYIDGLQDD 376

Qy    100 RII--------DPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYL 151
           ||         |  :   |      ::  |     || |     |              :
Db    377 AIIVLHSDGSHGPEYYERYPKEMERFQPVCRTNQLGSCSKEELVN-------------V 422

Qy    152 YDNNLINTNCVLTK 165
          ||| :: |:  |||
Db    423 YDNTILYTDHFLTK 436


RESULT   4
Q27167
ID   Q27167      PRELIMINARY;      PRT;  2395 AA.
AC   Q27167;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   51B TYPE SURFACE PROTEIN.
OS   Paramecium tetraurelia.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC   Paramecium.
OX   NCBI_TaxID=5888;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOCK 51;
RX   MEDLINE=95098630; PubMed=7800503;
RA   Scott J., Leeck C., Forney J.;
RT   "Analysis of the micronuclear B type surface protein gene in
RT   Paramecium tetraurelia.";
RL   Nucleic Acids Res. 22:5079-5084(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOCK 51;
RA   Scott J.M., Leeck C.L., Forney J.D.;
RL   Genetics 133:189-198(1993).
DR   EMBL; U07603; AAA81947.1; -.
DR   EMBL; L04795; AAA16710.1; -.
DR   InterPro; IPR002895; Paramecium_SA.
DR   InterPro; IPR003659; PSI.
```

```
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF01508; Paramecium_SA; 28.
DR   SMART; SM00423; PSI; 3.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
SQ   SEQUENCE   2395 AA;  246430 MW;  8632E022649B2D61 CRC64;


  Query Match             9.3%;  Score 92;  DB 5;  Length 2395;
  Best Local Similarity   22.4%;  Pred. No. 3.9;
  Matches   57;  Conservative   22;  Mismatches   76;  Indels  100;  Gaps   12;

Qy      10 TLNSLFNQIKNQSCGTSTASSPCITFRYPVD-----------GCYARAHKMRQILMNNG 57
           |:||      :::|  | | : |       |             || |   :
Db     487 TVNSTNAGCVDKTCENSLAQTIC-------DKDLNNKACIWKGKCYKRECVLASSTTATH 539

Qy      58 YDCEKQFVYGNLKASTGTCCV-------AWSYHVAILVSYKNASGV-------------- 96
            ||:   | |   ::||| ||       | :   |  :  :  |||
Db     540 ADCQTYDV-GCTLSNTGTGCVPLPLKCEAITIEAACNIRLQVTSGVKSYPLCGWNGSSCI 598

Qy      97 ---------TEKRIIDPSLFSSG-----------------PVTDTA-------------- 116
                   |     |   : ||                  | |  |
Db     599 DKACSTAPKTTATTSDCGTYKSGCVANNPVNGSIQGCQDLPTTCAARKSTENCEITRTGF 658

Qy     117 ----WR---NACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLL 169
               |    :|||  || :|||::   | | :   | :|   |   | |: |:    :
Db     659 PTCLWNSATSACVEKSCSTASVTT---TTGFLTVFSNTNCLAY---LSNSACIAN--NTA 710

Qy     170 SGCSPSPAPDVSSCG 184
           || | |    ||||
Db     711 DGCIPKP----SSCG 721


RESULT   5
Q9S746
ID   Q9S746      PRELIMINARY;      PRT;   594 AA.
AC   Q9S746;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ACE (AT1G72970/F3N23_17).
GN   ACE OR F3N23.17.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Araki T., Nakatani-Goto M.;
RT   "ADHESION OF CALYX EDGES (ACE), cDNA.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA   Altafi H., Araujo R., Huizar L., Rowley D., Chen S., Harman P.,
RA   Hicks R., Huerta M., Mason S., Siepel J., Zimmerman M., Buehler E.,
RA   Dunn P., Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J.,
RA   Liu S., Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA   Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA   Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA   Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA   Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA   Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA   Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB027458; BAA77837.1; -.
DR   EMBL; AC008017; AAD55644.1; -.
DR   EMBL; AY054193; AAL06854.1; -.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR000172; GMC_oxred.
DR   Pfam; PF00732; GMC_oxred; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00624; GMC_OXRED_2; UNKNOWN_1.
SQ   SEQUENCE   594 AA;  65344 MW;  D6A8682FBBF73881 CRC64;


  Query Match             9.2%;  Score 91;  DB 10;  Length 594;
  Best Local Similarity   24.6%;  Pred. No. 0.97;
  Matches   31;  Conservative   14;  Mismatches   45;  Indels   36;  Gaps    5;

Qy      25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMNNGY----DCEKQFVYGNLKAST-- 73
           |:   :| :|| |   |||   |      : ::: :| :     |:|| |:  |  |
Db     442 TNVDDNPSVTFNYFKHPVDLQRCVEAIRLVSKVVTSNRFLNYTQCDKQNVHKMLSLSVKA 501

Qy      74 --------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                              |    | ||   ||       :   ||   |:|| | |
Db     502 NINLRPKQLNDTKSMAQFCKDTVVTIWHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFD 561

Qy     109 SGPVTD 114
           | |:
Db     562 ESPGTN 567


RESULT   6
Q18804
ID   Q18804      PRELIMINARY;      PRT;   243 AA.
AC   Q18804;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 26.3 KDA PROTEIN.
GN   C53B7.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Du Z.;
RT   "The sequence of C. elegans cosmid C53B7.";
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U42830; AAC48274.1; -.
DR   HSSP; P02876; 9WGA.
KW   Hypothetical protein.
SQ   SEQUENCE   243 AA;  26299 MW;  6C1E33CB1CA2D16B CRC64;


  Query Match             9.1%;  Score 90;  DB 5;  Length 243;
  Best Local Similarity   26.7%;  Pred. No. 0.44;
  Matches   43;  Conservative   14;  Mismatches   64;  Indels   40;  Gaps    9;

Qy      22 SCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKAST--GTCCVA 79
           | | : ::  |::      | |:  :       || ||   |  |||   :|  || |
```

```
Db     39 SSGQTCSNGQCMSSLGYNGGLYSNTYN-----NNNQYDMNSQ--YGNQMGTTGSGTYCT- 90

Qy     80 WSYHVAILVSYKNASGVTEKRIIDPSLFS-----------SGPV----TDTAWRNACVN 123
                 ||   || |       | :|            ||      ||      | |
Db     91 --------TSYSCRSGETCMNNRCQSSYSNNMYNSNSQYSNSGSTCRYSTDCYSGQMCSN 142

Qy    124 TSCGSASVSSYANTAGNVYYRSPSNSY-LYDNNL-INTNCV 162
           |    |: |  :  |  | | : :|  |||   :|  ||
Db    143 GVC----VAQYGTSYNNAMYSSSTGNYCTYDNTCGLNQRCV 179


RESULT   7
O02364
ID   O02364      PRELIMINARY;      PRT;   927 AA.
AC   O02364; P90955; O62351;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 102.3 KDA PROTEIN T01D3.6 IN CHROMOSOME V.
GN   T01D3.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A
RC   STRAIN=BRISTOL N2;
RA   Steward C.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS; ISOFORM T01D3.6A (SHOWN HERE) AND
CC       ISOFORM T01D3.6B; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
DR   EMBL; Z81527; CAB04280.1; -.
DR   EMBL; Z81110; CAB04280.1; JOINED.
DR   EMBL; Z81110; CAB03262.1; -.
DR   EMBL; Z81527; CAB03262.1; JOINED.
DR   EMBL; Z81527; CAB04279.1; -.
DR   EMBL; Z81110; CAB04279.1; JOINED.
DR   EMBL; Z81110; CAB03263.1; -.
DR   EMBL; Z81527; CAB03263.1; JOINED.
DR   HSSP; P02671; 1FZD.
DR   WormPep; T01D3.6A; CE12964.
DR   WormPep; T01D3.6B; CE18164.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR002919; TIL.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00147; fibrinogen_C; 1.
DR   Pfam; PF01826; TIL; 1.
DR   Pfam; PF00094; vwd; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 2.
DR   SMART; SM00186; FBG; 1.
DR   SMART; SM00216; VWD; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Alternative splicing; Calcium-binding; EGF-like domain; Glycoprotein;
KW   Hydroxylation; Hypothetical protein; Repeat.
FT   VARSPLIC    405    416       MISSING (IN ISOFORM T01D3.6B).
SQ   SEQUENCE   927 AA;  103413 MW;  71AEF0A61FC2B266 CRC64;


  Query Match             9.0%;  Score 89.5;  DB 5;  Length 927;
  Best Local Similarity   20.1%;  Pred. No. 2.3;
  Matches   61;  Conservative   25;  Mismatches   68;  Indels  149;  Gaps   16;

Qy      4 VIPDVATLNSLFNQI--KNQSCGTSTASSPCITFRYPVDGCYARAHKMRQI------LMN 55
          | || |:     |:|  |: ||| :               ||   |::  |
Db    130 VCPDYASGRFCQNEIKCKDNSCGKNA------------DCYVANHQLNCICKPGYTARR 176

Qy     56 NGYDCEKQ------------FVYGNLKAS-TGTCCVAWS---------------------- 81
          || ||: :                 |  |:    |||   :|
Db    177 NGRDCDMKVQQACMSGDPHYVTYDGLRFDYQGTCPYVFSQPCTTLPAPYLWYSVRAKNEL 236

Qy     82 ----YHVA----ILVSYKN-------------ASGVTEKRIIDPSLFS-----------S 109
              ||::    : |   |              :||   ::: |  |             |
Db    237 PGKGYHISQVSEVEVDLHNLTIHVDGRSKTALVNGV---QVLTPWYFPNKNTWTVRVRFS 293

Qy    110 GP-----------VTDTAWRNACV-----------NTSCGSA------SVSSYANTAGNV 141
          |            || | : : ||           | || |       :    |  |:|
Db    294 GSTFTIENDQGVVVTFTTYNSLCVQVPDIPEFNGATTLCGLAGNIDGKKLDDVVNKNGSV 353

Qy    142 Y----YRSPSN-----------SYLYDNNLI----------------NTNCVLTKFSLLS 170
            | | |              ::: |  ||                |||||  |  ||
Db    354 LAIKSSRQPENNNHADFMKTEDTWITDKFLILRPGQENCINGQTLDNNTNCVSTSISLAQ 413

Qy    171 GCS 173
           |:
Db    414 SCA 416


RESULT   8
Q95QP7
ID   Q95QP7      PRELIMINARY;      PRT;   672 AA.
AC   Q95QP7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 71.4 KDA PROTEIN.
GN   D1044.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Pauley A.;
RT   "The sequence of C. elegans cosmid D1044.";
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U00065; AAL27237.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   672 AA;  71395 MW;  1717219E5C8375A2 CRC64;


  Query Match             9.0%;  Score 89;  DB 5;  Length 672;
  Best Local Similarity   24.1%;  Pred. No. 1.8;
  Matches   46;  Conservative   27;  Mismatches   88;  Indels   30;  Gaps    9;

Qy     20 NQSC----------GTSTASSPCITFRYPVDGCYARAHKM-RQILMNNGYDCEKQF--VY 66
          || |          || : :| |||   |  |  :  :  : ||   |  :  ||
Db    208 NQQCVISNGLNCQLGTVSYNSQCITLASPGQNCQTSSQCIDNSVCMNQMCTCNNNYRLVY 267

Qy     67 GNLKASTGTCCVAWSYHV---AILVSYKNASGVTEKRIIDPSLFSSGPVT----DTAWRN 119
          |     | : |      |    :|:|    : :  :: :  :: :||        ||
Db    268 GYCVPITSSICQQTQTLVNNQCVLLSIVGETCIANQQCVGGAMCNSGTCQCTNGATAMYG 327

Qy    120 ACVNTSCGSASVSSYANTAGNVYYR-----SPSNSYLYDNNLINTN--CVLTKFSLLSGC 172
           |:::|  | : |:  :  |  |       | | |    || : ||  || |  |:   |
```

```
Db    328 YCISSSSSSCN-SNQVSINGMCYNTVQVGGSCSFSQQCLNNAVCTNNICVSTFCSV--SC 384

Qy    173 SPSPAPDVSSC 183
          | :     : |
Db    385 STNQVCISNQC 395


RESULT   9
O26933
ID   O26933      PRELIMINARY;      PRT;   283 AA.
AC   O26933;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CONSERVED PROTEIN.
GN   MTH845.
OS   Methanothermobacter thermautotrophicus.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000861; AAB85343.1; -.
DR   InterPro; IPR002931; Transglut_core.
DR   Pfam; PF01841; Transglut_core; 1.
DR   SMART; SM00460; TGc; 1.
KW   Complete proteome.
SQ   SEQUENCE   283 AA;  32015 MW;  D89045B76AB76976 CRC64;


  Query Match             8.9%;  Score 88.5;  DB 17;  Length 283;
  Best Local Similarity   24.6%;  Pred. No. 0.73;
  Matches   29;  Conservative   17;  Mismatches   61;  Indels   11;  Gaps    2;

Qy     41 GCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGT-------CCVAWSYHVAILVSYKNA 93
          || :   |  :|        :  | |   | : ||       ||      |  :||:   |
Db    161 GCNSTYDKAVRIFNWVRDSIDYSFYYNTRKGAVGTLHSGSANCC----DHTHLLVALARA 216

Qy     94 SGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYL 151
          ||:  : :    :| ||      |    ||      |  :|::|  | |     |::::
Db    217 SGIPARYMHGNCVFRSGNTYGHVWGQLYVNGRWYDADATSFSNALGTVNNWDRSSAFI 274


RESULT  10
O94317
ID   O94317      PRELIMINARY;      PRT;   534 AA.
AC   O94317;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 53.4 KDA SERINE-RICH PROTEIN C215.13 IN CHROMOSOME II
DE   PRECURSOR.
GN   SPBC215.13.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- SIMILARITY: TO YEAST STA1.
DR   EMBL; AL033534; CAA22127.1; -.
KW   Hypothetical protein; Glycoprotein; Signal.
FT   SIGNAL        1     14       POTENTIAL.
FT   CHAIN        15    534       HYPOTHETICAL SERINE-RICH PROTEIN C215.13.
FT   CARBOHYD     31     31       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    426    426       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   534 AA;  53361 MW;  D62DC15C80A698E4 CRC64;


  Query Match             8.6%;  Score 85.5;  DB 3;  Length 534;
  Best Local Similarity   24.2%;  Pred. No. 3;
  Matches   45;  Conservative   32;  Mismatches   86;  Indels   23;  Gaps    5;

Qy      6 PDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFV 65
          |  ::  |:|::    |  |:: |||  ::  |             |:  :  :
Db     78 PTSSSEPSIFSESATPS-ETNSYSSPVSSYSDPATS---------QLPSSTSFFSPTSSE 127

Qy     66 YGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTS 125
          |      : :     |   ||| |    |   |  |   || || |:| : : :   :||
Db    128 YTPSSTESSSLLDPSSVSSAILPS----STSVEVSISSSSLSSSDPLTSSTFSSLSSSTS 183

Qy    126 CGSASVSS-YANTAGNVYYRSPSNSYLYDNNLI--------NTNCVLTKFSLLSGCSPSP 176
             ||||  ::|  :    | |:|||  ::::        :::  ||  || :   ||
Db    184 SSQPSVSSTSSSTFSSAAPTSTSSSYLSSSSVVSSSSSPSSSSSSTLTSSSLSTSSIPST 243

Qy    177 APDVSS 182
          :  ||
Db    244 SSSSSS 249


RESULT  11
Q9SXZ3
ID   Q9SXZ3      PRELIMINARY;      PRT;   594 AA.
AC   Q9SXZ3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ACE.
GN   ACE.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NOSSEN;
RA   Araki T., Nakatani-Goto M.;
RT   "Arabidopsis ADHESION OF CALYX EDGES (ACE), genomic.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB027507; BAA77842.1; -.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR000172; GMC_oxred.
DR   Pfam; PF00732; GMC_oxred; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00624; GMC_OXRED_2; UNKNOWN_1.
SQ   SEQUENCE   594 AA;  65358 MW;  26A86D2FB82C53EC CRC64;


  Query Match             8.6%;  Score 85;  DB 10;  Length 594;
  Best Local Similarity   24.6%;  Pred. No. 3.8;
  Matches   31;  Conservative   13;  Mismatches   46;  Indels   36;  Gaps    5;

Qy     25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMN----NGYDCEKQFVYGNLKAST-- 73
          |:   :| :|| |   |||   |       : ::: :    |   |:|| |:  |  |
Db    442 TNVDDNPSVTFNYFKHPVDLQRCVEAIRLVSKVVTSKRFLNYTQCDKQNVHKMLSLSVKA 501

Qy     74 --------------------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFS 108
                              |    | ||   ||       :   ||   |:|| | |
```

```
Db    502 NINLRPKQLNDTKSMAQFCKDTVVTIWHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFD 561

Qy    109 SGPVTD 114
          | |:
Db    562 ESPGTN 567


RESULT  12
P70004
ID   P70004      PRELIMINARY;      PRT;   342 AA.
AC   P70004;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ALPHA1-MICROGLOBULIN/BIKUNIN (AMBP) PRECURSOR.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Kawahara A., Hikosaka A., Sasado T., Hirota K.;
RT   "Thyroid hormone-dependent repression of alphal-microglobulin/bikunin
RT   (AMBP) gene expression during amphibian metamorphosis.";
RL   Dev. Genes Evol. 206:355-362(1997).
DR   EMBL; D87752; BAA13453.1; -.
DR   HSSP; P02760; 1BIK.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   InterPro; IPR002345; Lipocalin.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF00014; Kunitz_BPTI; 2.
DR   Pfam; PF00061; lipocalin; 1.
DR   PRINTS; PR00759; BASICPTASE.
DR   PRINTS; PR00179; LIPOCALIN.
DR   SMART; SM00131; KU; 2.
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 2.
DR   PROSITE; PS00213; LIPOCALIN; 1.
KW   Serine protease inhibitor; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    342       POTENTIAL.
SQ   SEQUENCE   342 AA;  38558 MW;  3050508BFDC5C47B CRC64;


  Query Match             8.5%;  Score 84;  DB 13;  Length 342;
  Best Local Similarity   25.2%;  Pred. No. 2.6;
  Matches   31;  Conservative   14;  Mismatches   32;  Indels   46;  Gaps    7;

Qy     12 NSLFNQIKNQSCGTSTASSPCI----------------TFRYPVDGCYARAHKMRQILMN 55
          || |:: | :||  : || ||:                ||:|   ||           |
Db    214 NSPFSKNKGESCRLAPASGPCLGNHNRYFYNSSTMACETFQY--GGCLGN---------N 262

Qy     56 NGYDCEKQFVYG-------NLKASTGTCCVA---WSYHVA----ILVSYKNASG-----V 96
          | :  ||: ::         |  : | |  |   |::  |    :  ||    |
Db    263 NNFHSEKECLHDCRTEAACRLPITPGPCKTAKTHWAFDAAQGKCVTFSYGGCQGNGNQFY 322

Qy     97 TEK 99
          |||
Db    323 TEK 325


RESULT  13
Q9VFL1
ID   Q9VFL1      PRELIMINARY;      PRT;  3726 AA.
AC   Q9VFL1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TRX PROTEIN.
GN   TRX OR CG8651.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F.,. Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR   EMBL; AE003704; AAF55041.2; -.
DR   HSSP; P20393; 1A6Y.
DR   FlyBase; FBgn0003862; trx.
DR   InterPro; IPR003889; FYrich_C.
DR   InterPro; IPR003888; FYrich_N.
DR   InterPro; IPR001965; PHD.
DR   InterPro; IPR003616; PostSET.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00628; PHD; 2.
DR   Pfam; PF00856; SET; 1.
DR   SMART; SM00542; FYRC; 1.
DR   SMART; SM00541; FYRN; 1.
DR   SMART; SM00249; PHD; 4.
DR   SMART; SM00508; PostSET; 1.
DR   SMART; SM00184; RING; 3.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS50280; SET; 1.
SQ   SEQUENCE   3726 AA;  400095 MW;  E3DDB8F062BD7796 CRC64;


  Query Match             8.5%;  Score 84;  DB 5;  Length 3726;
  Best Local Similarity   25.3%;  Pred. No. 40;
  Matches   46;  Conservative   21;  Mismatches   73;  Indels   42;  Gaps    7;

Qy     22 SCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEK------QFVYGNLKAST-- 73
```

```
          | | |: |||        ||            ::|  | |      :   | |  :|
Db    178 SSGRSSGSSPDGNSGASSDGA-----------SSGISCGKSTAKSTEASSGKLAKTTGA 225

Qy     74 GTCCVAWSYHVAILVSYKNASGVTEKRIID-------PSLFSSGPVTDTAWRNACVNTSC 126
          ||| |         | | :|| | :           :|| : | | |    | |:
Db    226 GTCSSA--------KSSKASSGTTSEATTSGLSGACLKALFVATPATSTGLACALVSPG- 276

Qy    127 GSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGC------SPSPAPDV 180
          ||:   ::  :|  :  |  ||    : ||     :|   | |         :|||:| :
Db    277 GSSQGGTFPISAALLRARKNSNKKFKNLNLARGEVMLPSTSKLKQLNSPVVDNPSPSPPI 336

Qy    181 SS 182
          :|
Db    337 AS 338


RESULT  14
Q9GTE0
ID   Q9GTE0      PRELIMINARY;      PRT;   619 AA.
AC   Q9GTE0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NUCLEAR RECEPTOR NHR-66 (FRAGMENT).
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bogan A., Maina C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT   "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
RT   compatibility with the ligand-binding domain fold.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF273819; AAG15168.1; -.
DR   HSSP; P20393; 1A6Y.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   619 AA;  69082 MW;  6580E456F0E46D48 CRC64;


  Query Match             8.4%;  Score 83.5;  DB 5;  Length 619;
  Best Local Similarity   21.5%;  Pred. No. 5.7;
  Matches   49;  Conservative   20;  Mismatches   66;  Indels   93;  Gaps   11;

Qy     23 CGTSTASSPCITFRYPVDGCYA-RAHKMRQILMNNGYDCEKQFVYGNLK-ASTGTCCVAW 80
          ||| :         : || | |  |   | :::|  | | |      :|  | |  | |
Db    158 CGTDSTG-----IHFGVDACAACSAFFRRTVVLNKDYSCNKGGKCTVVKDGSAGQKCRAC 212

Qy     81 SYHVAILVSY-------------KNASGVTEKRIIDPSL--------------FSSGP- 111
           :   |                 | ::||  || : |                  ||||
Db    213 RFRKCISSGMDKNSVQHRRDAIGKYSAGV--KRELSPDAEFEPSAKVSTVSEPSTSSGPS 270

Qy    112 ------------------VTDTAWRNACVNTSCGSASVSS---------------YANT 137
                            |   : |:|::||  ||                   ||
Db    271 GGFNQNVSSPAGIPRVPSTLRTTQASTCMNSACGQKSVLHELICRQNFLTEQRQLFYAGC 330

Qy    138 AGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCGF 185
           |: ::| ||:    :|  :                 |   | ||| |
Db    331 LGD-WFRKPSS---IENQTL----------------SELTDFSSCMF 357


RESULT  15
Q9GTD9
ID   Q9GTD9      PRELIMINARY;      PRT;   624 AA.
AC   Q9GTD9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NUCLEAR RECEPTOR NHR-66 (FRAGMENT).
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bogan A., Maina C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT   "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
RT   compatibility with the ligand-binding domain fold.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF273820; AAG15169.1; -.
DR   HSSP; P20393; 1A6Y.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   624 AA;  69495 MW;  DC5D1DEE608ADAF0 CRC64;


  Query Match             8.4%;  Score 83.5;  DB 5;  Length 624;
  Best Local Similarity   21.5%;  Pred. No. 5.7;
  Matches   49;  Conservative   20;  Mismatches   66;  Indels   93;  Gaps   11;

Qy     23 CGTSTASSPCITFRYPVDGCYA-RAHKMRQILMNNGYDCEKQFVYGNLK-ASTGTCCVAW 80
          ||| :         : || | |  |   | :::|  | | |      :|  | |  | |
Db    163 CGTDSTG-----IHFGVDACAACSAFFRRTVVLNKDYSCNKGGKCTVVKDGSAGQKCRAC 217

Qy     81 SYHVAILVSY-------------KNASGVTEKRIIDPSL--------------FSSGP- 111
           :   |                 | ::||  || : |                  ||||
Db    218 RFRKCISSGMDKNSVQHRRDAIGKYSAGV--KRELSPDAEFEPSAKVSTVSEPSTSSGPS 275

Qy    112 ------------------VTDTAWRNACVNTSCGSASVSS---------------YANT 137
                            |   : |:|::||  ||                   ||
Db    276 GGFNQNVSSPAGIPRVPSTLRTTQASTCMNSACGQKSVLHELICRQNFLTEQRQLFYAGC 335

Qy    138 AGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCGF 185
           |: ::| ||:    :|  :                 |   | ||| |
Db    336 LGD-WFRKPSS---IENQTL----------------SELTDFSSCMF 362


Search completed: June 27, 2002, 21:58:26
Job time: 659 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 20:16:15 ; Search time 3055.3 Seconds
(without alignments)
7397.198 Million cell updates/sec

Title: US-09-727-769A-7
Perfect score: 1080
Sequence: 1 agttaaaataaccaaccaac.....atcagcaggtgctgttatat 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	1080	100.0	1080	6	AX113623	AX113623 Sequence
2	1080	100.0	1380	1	AB046594	AB046594 Chryseoba
3	555	51.4	555	6	AX113621	AX113621 Sequence
4	527.4	48.8	1080	6	AR159962	AR159962 Sequence
5	527.4	48.8	1080	6	E59342	E59342 Novel prote
6	344.2	31.9	555	6	AR159957	AR159957 Sequence
7	344.2	31.9	555	6	E59337	E59337 Novel prote
c 8	60	5.6	7218	6	I66494	I66494 Sequence 14
c 9	45.4	4.2	7319	6	AX346947	AX346947 Sequence
10	44.4	4.1	5993	8	AB018378	AB018378 Glycine m
11	44.4	4.1	13776	1	STATRSC	L11998 Staphylococ
12	44.4	4.1	46445	1	AF051917	AF051917 Staphyloc
13	44.4	4.1	197405	8	ATCHRIV40	AL161540 Arabidops
14	44.4	4.1	202860	8	ATFCA2	Z97337 Arabidopsis
c 15	44	4.1	1182	9	BC016000	BC016000 Homo sapi
c 16	43.6	4.0	20992	3	LEIKPMAX	M10126 Leishmania
c 17	43.6	4.0	43049	9	HSDJ87H18	AL121846 Human DNA
c 18	43.6	4.0	117445	2	AL160263	AL160263 Homo sapi
c 19	43.4	4.0	2324	5	XELNRP1B	M34895 X.laevis ne
c 20	43.2	4.0	92633	2	PFMAL4P1_3	Continuation (4 of
c 21	43	4.0	145640	2	AC018361	AC018361 Homo sapi
22	43	4.0	158254	2	AC098659	AC098659 Rattus no
23	42.8	4.0	1141	6	AX083744	AX083744 Sequence
c 24	42.8	4.0	1141	6	AX083744	AX083744 Sequence
c 25	42.8	4.0	122051	8	AC005970	AC005970 Arabidops
c 26	42.6	3.9	88064	2	AC022707	AC022707 Homo sapi
c 27	42.4	3.9	17131	6	AX345955	AX345955 Sequence
28	42.4	3.9	219244	2	AL670771	AL670771 Mus muscu
c 29	42.2	3.9	6314	6	AX281272	AX281272 Sequence
30	42	3.9	125020	9	AF429315	AF429315 Homo sapi
31	41.8	3.9	1857	6	AR157645	AR157645 Sequence
32	41.8	3.9	77835	2	PFMAL13P2_3	Continuation (4 of
33	41.8	3.9	181596	2	AC068158	AC068158 Homo sapi
c 34	41.6	3.9	5987	6	AX346465	AX346465 Sequence
c 35	41.4	3.8	5884	6	AX347067	AX347067 Sequence
36	41.4	3.8	55984	3	U97196	U97196 Caenorhabdi
37	41.4	3.8	56511	3	AC084467	AC084467 Caenorhab
38	41.2	3.8	62423	2	AP004554	AP004554 Oryza sat
39	41.2	3.8	82812	9	HSJ764O22	AL096799 Human DNA
c 40	41.2	3.8	102237	2	AC099322	AC099322 Danio rer
c 41	41	3.8	7508	6	AX344203	AX344203 Sequence
42	41	3.8	73345	8	ALO294725	AJ294725 Astasia l
43	41	3.8	84563	5	AC096885	AC096885 Danio rer
c 44	41	3.8	125020	9	AF429315	AF429315 Homo sapi
c 45	41	3.8	136198	2	AL392004	AL392004 Homo sapi

ALIGNMENTS

RESULT 1
AX113623
LOCUS AX113623 1080 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 7 from Patent EP1106696.
ACCESSION AX113623
VERSION AX113623.1 GI:13939800
KEYWORDS .
SOURCE Chryseobacterium sp. No. 9670.
ORGANISM Chryseobacterium sp. No. 9670
Bacteria; CFB group; Flavobacteria; Flavobacteriaceae;
Chryseobacterium.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Yamaguchi,S.
TITLE Protein-deamidating enzyme, microorganism producing the same, gene encoding the same, production process therefor, and use thereof
JOURNAL Patent: EP 1106696-A 7 13-JUN-2001;
Amano Enzyme Inc. (JP)
FEATURES Location/Qualifiers
source 1. .1080
/organism="Chryseobacterium sp. No. 9670"
/db_xref="taxon:161487"
CDS 61. .1023
/note="unnamed protein product"

```
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC37857.1"
                     /db_xref="GI:13939801"
                     /translation="MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKD
                     FGKTVPVGIDEENGMIKVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKP
                     NSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGT
                     STASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYH
                     VAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNV
                     YYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCGF"
     mat_peptide     466. .1020
                     /product="unnamed"
BASE COUNT      371 a    204 c    203 g    302 t
ORIGIN


  Query Match           100.0%;  Score 1080;  DB 6;  Length 1080;
  Best Local Similarity  100.0%;  Pred. No. 7.2e-236;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 agttaaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattatt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGTTAAAATAACCAACCAACTTAACAAAAACTCACCATTAAACTACAAATTACAATTATT 60

Qy     61 atgaaaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATGAAAAATCTTTTTTTATCAATGATGGCCTTTGTGACCGTCTTAACTTTTAATTCCTGT 120

Qy    121 gccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattct 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCGATTCCAACGGGAATCAGGAAATCAACGGAAAGGAAAAACTAAGTGTAAATGATTCT 180

Qy    181 aagctgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgata 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAGCTGAAAGATTTCGGAAAGACTGTACCGGTAGGGATAGACGAAGAAAACGGAATGATA 240

Qy    241 aaggtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgag 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAGGTGTCATTTATGTTAACTGCGCAATTCTATGAAATTAAGCCGACCAAAGAAAATGAG 300

Qy    301 cagtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttctta 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGTATATCGGAATGCTTAGACAGGCTGTTAAGAATGAATCTCCTGTACACATTTTCTTA 360

Qy    361 aagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGCCTAATAGCAATGAAATAGGAAAAGTGGAGTCTGCAAGTCCGGAAGACGTAAGATAT 420

Qy    421 tttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTTAAAACGATCCTGACAAAAGAAGTAAAAGGGCAAACCAATAAATTGGCGAGTGTAATT 480

Qy    481 cctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttgcggtacc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCTGATGTAGCTACATTAAATTCTTTATTCAATCAAATAAAGAATCAGTCTTGCGGTACC 540

Qy    541 tctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgttatgcaaga 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCTACGGCGTCCTCACCATGCATCACATTCAGATATCCTGTAGACGGATGTTATGCAAGA 600

Qy    601 gcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttgta 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCCCATAAGATGAGACAAATCTTAATGAACAACGGCTATGACTGTGAAAAACAATTTGTA 660

Qy    661 tacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctaccacgttgca 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TACGGAAACCTAAAGGCATCAACAGGAACTTGCTGTGTGGCGTGGAGCTACCACGTTGCA 720

Qy    721 atattggtaagctataaaaatgcttccggagtaacggaaaaaagaattattgatccttca 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ATATTGGTAAGCTATAAAAATGCTTCCGGAGTAACGGAAAAAAGAATTATTGATCCTTCA 780

Qy    781 ctattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgttaacacctct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTATTTTCAAGCGGTCCTGTAACAGATACAGCATGGAGAAACGCTTGCGTTAACACCTCT 840

Qy    841 tgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttattacagaagt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TGCGGATCTGCATCCGTTTCCTCTTATGCTAATACTGCAGGAAATGTTTATTACAGAAGT 900

Qy    901 cctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgtactgactaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCTAGTAATTCTTACCTGTATGACAACAATCTGATCAATACCAACTGTGTACTGACTAAA 960

Qy    961 ttttcactgctttccggatgttctccttcacctgcaccggatgtatccagctgtggattt 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTTTCACTGCTTTCCGGATGTTCTCCTTCACCTGCACCGGATGTATCCAGCTGTGGATTT 1020

Qy   1021 taattaattgataattttacagcacctgctcatttacagaatcagcaggtgctgttatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TAATTAATTGATAATTTTACAGCACCTGCTCATTTACAGAATCAGCAGGTGCTGTTATAT 1080


RESULT   2
AB046594
LOCUS       AB046594                1380 bp    DNA     linear   BCT 27-APR-2001
DEFINITION  Chryseobacterium proteolyticum prgA gene for protein-glutaminase,
            complete cds.
ACCESSION   AB046594
VERSION     AB046594.1  GI:12597204
KEYWORDS    .
SOURCE      Chryseobacterium proteolyticum (strain:9670) DNA.
  ORGANISM  Chryseobacterium proteolyticum
            Bacteria; CFB group; Flavobacteria; Flavobacteriaceae;
            Chryseobacterium.
REFERENCE   1  (sites)
  AUTHORS   Yamaguchi,S., Jeenes,D.J. and Archer,D.B.
  TITLE     Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme
            that deamidates glutaminyl residues in proteins Purification,
            characterization and gene cloning
  JOURNAL   Eur. J. Biochem. 268 (5), 1410-1421 (2001)
  MEDLINE   21153247
REFERENCE   2  (bases 1 to 1380)
  AUTHORS   Yamaguchi,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Shotaro Yamaguchi, Amano Pharmaceutical Co.
            Ltd., Research and Development; Suei, Kagamigahara, Gifu 509-0108,
            Japan (E-mail:LDV01447@nifty.ne.jp, Tel:81-583-79-1220,
            Fax:81-583-79-1232)
FEATURES             Location/Qualifiers
     source          1. .1380
                     /organism="Chryseobacterium proteolyticum"
                     /strain="9670"
                     /db_xref="taxon:118127"
     sig_peptide     181. .243
                     /gene="prgA"
     CDS             181. .1143
                     /gene="prgA"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-glutaminase"
                     /protein_id="BAB21508.1"
                     /db_xref="GI:12597205"
                     /translation="MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKD
                     FGKTVPVGIDEENGMIKVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKP
                     NSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGT
                     STASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYH
                     VAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNV
                     YYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCGF"
     gene            181. .1143
                     /gene="prgA"
     mat_peptide     586. .1140
                     /gene="prgA"
                     /evidence=experimental
                     /product="unnamed"
BASE COUNT      472 a    247 c    247 g    414 t
```

```
ORIGIN

  Query Match             100.0%;  Score 1080;  DB 1;  Length 1380;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-236;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 agttaaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattatt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGTTAAAATAACCAACCAACTTAACAAAAACTCACCATTAAACTACAAATTACAATTATT 180

Qy      61 atgaaaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATGAAAAATCTTTTTTTATCAATGATGGCCTTTGTGACCGTCTTAACTTTTAATTCCTGT 240

Qy     121 gccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattct 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCCGATTCCAACGGGAATCAGGAAATCAACGGAAAGGAAAAACTAAGTGTAAATGATTCT 300

Qy     181 aagctgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgata 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAGCTGAAAGATTTCGGAAAGACTGTACCGGTAGGGATAGACGAAGAAAACGGAATGATA 360

Qy     241 aaggtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgag 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAGGTGTCATTTATGTTAACTGCGCAATTCTATGAAATTAAGCCGACCAAAGAAAATGAG 420

Qy     301 cagtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttctta 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAGTATATCGGAATGCTTAGACAGGCTGTTAAGAATGAATCTCCTGTACACATTTTCTTA 480

Qy     361 aagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AAGCCTAATAGCAATGAAATAGGAAAAGTGGAGTCTGCAAGTCCGGAAGACGTAAGATAT 540

Qy     421 tttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaatt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTTAAAACGATCCTGACAAAAGAAGTAAAAGGGCAAACCAATAAATTGGCGAGTGTAATT 600

Qy     481 cctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttgcggtacc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CCTGATGTAGCTACATTAAATTCTTTATTCAATCAAATAAAGAATCAGTCTTGCGGTACC 660

Qy     541 tctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgttatgcaaga 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TCTACGGCGTCCTCACCATGCATCACATTCAGATATCCTGTAGACGGATGTTATGCAAGA 720

Qy     601 gcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttgta 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCCCATAAGATGAGACAAATCTTAATGAACAACGGCTATGACTGTGAAAAACAATTTGTA 780

Qy     661 tacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctaccacgttgca 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TACGGAAACCTAAAGGCATCAACAGGAACTTGCTGTGTGGCGTGGAGCTACCACGTTGCA 840

Qy     721 atattggtaagctataaaaatgcttccggagtaacggaaaaaagaattattgatccttca 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATATTGGTAAGCTATAAAAATGCTTCCGGAGTAACGGAAAAAAGAATTATTGATCCTTCA 900

Qy     781 ctattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgttaacacctct 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CTATTTTCAAGCGGTCCTGTAACAGATACAGCATGGAGAAACGCTTGCGTTAACACCTCT 960

Qy     841 tgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttattacagaagt 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TGCGGATCTGCATCCGTTTCCTCTTATGCTAATACTGCAGGAAATGTTTATTACAGAAGT 1020

Qy     901 cctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgtactgactaaa 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCTAGTAATTCTTACCTGTATGACAACAATCTGATCAATACCAACTGTGTACTGACTAAA 1080

Qy     961 ttttcactgctttccggatgttctccttcacctgcaccggatgtatccagctgtggattt 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TTTTCACTGCTTTCCGGATGTTCTCCTTCACCTGCACCGGATGTATCCAGCTGTGGATTT 1140

Qy    1021 taattaattgataattttacagcacctgctcatttacagaatcagcaggtgctgttatat 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TAATTAATTGATAATTTTACAGCACCTGCTCATTTACAGAATCAGCAGGTGCTGTTATAT 1200


RESULT   3
AX113621
LOCUS       AX113621                555 bp    DNA     linear   PAT 30-MAY-2001
DEFINITION  Sequence 5 from Patent EP1106696.
ACCESSION   AX113621
VERSION     AX113621.1  GI:13939799
KEYWORDS    .
SOURCE      Chryseobacterium sp. No. 9670.
  ORGANISM  Chryseobacterium sp. No. 9670
            Bacteria; CFB group; Flavobacteria; Flavobacteriaceae;
            Chryseobacterium.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Yamaguchi,S.
  TITLE     Protein-deamidating enzyme, microorganism producing the same, gene
            encoding the same, production process therefor, and use thereof
  JOURNAL   Patent: EP 1106696-A 5 13-JUN-2001;
            Amano Enzyme Inc. (JP)
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Chryseobacterium sp. No. 9670"
                     /db_xref="taxon:161487"
BASE COUNT      166 a    119 c    107 g    163 t
ORIGIN

  Query Match             51.4%;  Score 555;  DB 6;  Length 555;
  Best Local Similarity   100.0%;  Pred. No. 3e-116;
  Matches  555;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     466 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 525
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTGGCGAGTGTAATTCCTGATGTAGCTACATTAAATTCTTTATTCAATCAAATAAAGAAT 60

Qy     526 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 585
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CAGTCTTGCGGTACCTCTACGGCGTCCTCACCATGCATCACATTCAGATATCCTGTAGAC 120

Qy     586 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 645
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGATGTTATGCAAGAGCCCATAAGATGAGACAAATCTTAATGAACAACGGCTATGACTGT 180

Qy     646 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 705
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAAAAACAATTTGTATACGGAAACCTAAAGGCATCAACAGGAACTTGCTGTGTGGCGTGG 240

Qy     706 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 765
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGCTACCACGTTGCAATATTGGTAAGCTATAAAAATGCTTCCGGAGTAACGGAAAAAAGA 300

Qy     766 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 825
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATTATTGATCCTTCACTATTTTCAAGCGGTCCTGTAACAGATACAGCATGGAGAAACGCT 360

Qy     826 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 885
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGCGTTAACACCTCTTGCGGATCTGCATCCGTTTCCTCTTATGCTAATACTGCAGGAAAT 420

Qy     886 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 945
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GTTTATTACAGAAGTCCTAGTAATTCTTACCTGTATGACAACAATCTGATCAATACCAAC 480

Qy     946 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 1005
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TGTGTACTGACTAAATTTTCACTGCTTTCCGGATGTTCTCCTTCACCTGCACCGGATGTA 540

Qy    1006 tccagctgtggattt 1020
```

```
Db    541 TCCAGCTGTGGATTT 555


RESULT   4
AR159962
LOCUS       AR159962               1080 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 11 from patent US 6251651.
ACCESSION   AR159962
VERSION     AR159962.1  GI:16222841
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Yamaguchi,S. and Matsuura,A.
  TITLE     Protein-deamidating enzyme, gene encoding the same, production
            process therefor, and use thereof
  JOURNAL   Patent: US 6251651-A 11 26-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="unknown"
BASE COUNT      354 a    210 c    205 g    311 t
ORIGIN


  Query Match           48.8%;  Score 527.4;  DB 6;  Length 1080;
  Best Local Similarity   70.9%;  Pred. No. 5.5e-110;
  Matches  715;  Conservative    0;  Mismatches  291;  Indels    3;  Gaps    1;

Qy     17 caacttaacaaaaactcaccattaaactacaaattacaattattatgaaaaatctttttt 76
          ||| | || |       |  | |||||  |||| || ||  | |||||||||  || | |
Db     17 CAATTAAAAAGTTCACTAAAACTAAACACCAAAATATAAAAACTATGAAAAAATTTCTGT 76

Qy     77 tatcaatgatggccctttgtgaccgtcttaacttttaattcctgtgccgattccaacggga 136
          |||| |||||||| || |||||  |  |  | || ||| ||||  | |||||| |  |  |
Db     77 TATCCATGATGGCATTCGTGACGATTCTGTCATTCAATGCCTGCTCAGATTCAAGTGCCA 136

Qy    137 atcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagatttcg 196
          | ||||||    ||       |  ||| |  |    | |   |||   ||||||||||||
Db    137 ACCAGGACCCGAATCTTGTCGCTAAAGAATCTAACGAAGTCGCTA---TGAAAGATTTCG 193

Qy    197 gaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatttatgt 256
          | ||||||| |||||||||| ||  ||||| | ||||   | ||  | |||||||||
Db    194 GTAAGACTGTTCCGGTAGGGATTGAAAAAGAAGATGGAAAATTTAAAATCTCATTTATGG 253

Qy    257 taactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaatgc 316
          | ||||| ||    |||||||||  |   |  ||||||||||    |||||   |||
Db    254 TTACTGCCCAGCCGTATGAAATTGCGGACAGTAAAGAAAATGCAGGTTATATTTCCATGA 313

Qy    317 ttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaatg 376
          | |||||||||||| |||||||| |||| || ||  ||||| | ||    || | ||||
Db    314 TCAGACAGGCTGTTGAGAATGAAACTCCCGTTCATGTTTTCCTTAAAGTCAACACCAATA 373

Qy    377 aaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcctga 436
          |||| | |||||| ||    ||||     || ||| |  | |||||||||  |  |  | |
Db    374 AAATTGCAAAAGTAGAAAAAGCAACAGATGATGACATCCGTTATTTTAAATCTGTATTCA 433

Qy    437 caaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctgatgtagctacat 496
            ||  |||  | |||  ||| ||| |||   |  ||||| |||||||||  | || ||
Db    434 ACAAGCAAGAGAGAGGTGAAAGCAACAAAGCAGTCAGTGTTATTCCTGATCTGGCAACGC 493

Qy    497 taaattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctcac 556
          | ||   |||||| |  || || || || ||| ||||||| || ||||| || || || |
Db    494 TGAACAGTTTATTTACCCAGATCAAAAACCAGGCTTGCGGAACTTCTACAGCATCTTCTC 553

Qy    557 catgcatcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagac 616
          | || ||||| |||||||||| || ||||||||||||||||| || || || |||||||
Db    554 CTTGTATCACCTTCAGATATCCGGTTGACGGATGTTATGCAAGGGCTCACAAAATGAGAC 613

Qy    617 aaatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaagg 676
          ||||| || |||||  |||||||||||||||||| || || ||||| || || || |  |
Db    614 AAATCCTATTGAACGCCGGCTATGACTGTGAAAAGCAGTTCGTATATGGTAATCTGAGAG 673

Qy    677 catcaacaggaacttgctgtgtggcgtggagctaccacgttgcaatattggtaagctata 736
          | || |||||||| ||||||||  | |||   || ||||| ||||| ||||||||||  |
Db    674 CTTCTACAGGAACATGCTGTGTATCATGGGTATATCACGTAGCAATTTTGGTAAGCTTCA 733

Qy    737 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 796
          |||||||||| ||| |    |||||||||| || ||||||||| |||| || |||||||
Db    734 AAAATGCTTCAGGAATTGTTGAAAAAAGAATCATAGATCCTTCATTATTCTCCAGCGGTC 793

Qy    797 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 856
          |||||||||| | |||||||||   || ||    |||||    |||||||||| || |
Db    794 CTGTAACAGATTCTGCATGGAGAGCTGCATGTACCAACACAAGCTGCGGATCTGCGTCTG 853

Qy    857 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaatttcttacc 916
          | || || || || ||||| |||||||||||||| ||||||||||      ||| |  |
Db    854 TATCTTCCTACGCCAATACAGCAGGAAATGTTTACTACAGAAGTCCGTCAGGTTCATTAC 913

Qy    917 tgtatgacaacaatctgatcaataccaactgtgtactgactaaatttttcactgctttccg 976
          ||||||| |||||     | |||||||| |||||| | |   | ||| |||   ||||| |
Db    914 TGTATGATAACAACTATGTGAATACCAATTGTGTATTAAACATATTCTCATCCCTTTCAG 973

Qy    977 gatgttctccttcacctgcaccggatgtatccagctgtggatttttaatt 1025
          |||||||||||| || |||||   |||| | ||||||||||||||||||
Db    974 GATGTTCTCCTTCCCCAGCACCAAGTGTAGCAAGCTGTGGATTTTAATT 1022


RESULT   5
E59342
LOCUS       E59342                 1080 bp    DNA     linear   PAT 07-FEB-2001
DEFINITION  Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof.
ACCESSION   E59342
VERSION     E59342.1  GI:13023309
KEYWORDS    JP 2000050887-A/8.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Shotaro,Y.A.M.M.
  TITLE     Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof
  JOURNAL   Patent: JP 2000050887-A 8 22-FEB-2000;
            AMANO PHARMACEUT CO LTD
COMMENT     OS   Chryseobacterium gleum
            PN   JP 2000050887-A/8
            PD   22-FEB-2000
            PF   04-JUN-1999 JP 1999158703
            PR
            PI   SHOTARO YAMAGUCHI,AKIRA MATSUURA
            PC   C12N15/09,C12N9/80//A21D2/36,A23C9/154,A23C11/06,A23J3/00, PC
             A23J3/10,
            PC   A23J3/16,A23J3/18,A23J3/34,A23L1/176,A23L1/23,A23L1/317, PC
            (C12N9/80,C12R1:01),(C12N9/80,C12R1:20),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (61). .(1020)
            FT   mat_peptide     (463). .(1017).
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      354 a    210 c    205 g    311 t
ORIGIN


  Query Match           48.8%;  Score 527.4;  DB 6;  Length 1080;
  Best Local Similarity   70.9%;  Pred. No. 5.5e-110;
  Matches  715;  Conservative    0;  Mismatches  291;  Indels    3;  Gaps    1;

Qy     17 caacttaacaaaaactcaccattaaactacaaattacaattattatgaaaaatctttttt 76
          ||| | || |       |  | |||||  |||| || ||  | |||||||||  || | |
Db     17 CAATTAAAAAGTTCACTAAAACTAAACACCAAAATATAAAAACTATGAAAAAATTTCTGT 76

Qy     77 tatcaatgatggccctttgtgaccgtcttaacttttaattcctgtgccgattccaacggga 136
```

```
Db     77 TATCCATGATGGCATTCGTGACGATTCTGTCATTCAATGCCTGCTCAGATTCAAGTGCCA 136

Qy    137 atcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagatttcg 196
Db    137 ACCAGGACCCGAATCTTGTCGCTAAAGAATCTAACGAAGTCGCTA---TGAAAGATTTCG 193

Qy    197 gaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatttatgt 256
Db    194 GTAAGACTGTTCCGGTAGGGATTGAAAAAGAAGATGGAAAATTTAAAATCTCATTTATGG 253

Qy    257 taactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaatgc 316
Db    254 TTACTGCCCAGCCGTATGAAATTGCGGACAGTAAAGAAAATGCAGGTTATATTTCCATGA 313

Qy    317 ttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaatg 376
Db    314 TCAGACAGGCTGTTGAGAATGAAACTCCCGTTCATGTTTTCCTTAAAGTCAACACCAATA 373

Qy    377 aaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcctga 436
Db    374 AAATTGCAAAAGTAGAAAAAGCAACAGATGATGACATCCGTTATTTTAAATCTGTATTCA 433

Qy    437 caaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctgatgtagctacat 496
Db    434 ACAAGCAAGAGAGAGGTGAAAGCAACAAAGCAGTCAGTGTTATTCCTGATCTGGCAACGC 493

Qy    497 taaattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctcac 556
Db    494 TGAACAGTTTATTTACCCAGATCAAAAACCAGGCTTGCGGAACTTCTACAGCATCTTCTC 553

Qy    557 catgcatcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagac 616
Db    554 CTTGTATCACCTTCAGATATCCGGTTGACGGATGTTATGCAAGGGCTCACAAAATGAGAC 613

Qy    617 aaatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaagg 676
Db    614 AAATCCTATTGAACGCCGGCTATGACTGTGAAAAGCAGTTCGTATATGGTAATCTGAGAG 673

Qy    677 catcaacaggaacttgctgtgtggcgtggagctaccacgttgcaatattggtaagctata 736
Db    674 CTTCTACAGGAACATGCTGTGTATCATGGGTATATCACGTAGCAATTTTGGTAAGCTTCA 733

Qy    737 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 796
Db    734 AAAATGCTTCAGGAATTGTTGAAAAAAGAATCATAGATCCTTCATTATTCTCCAGCGGTC 793

Qy    797 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 856
Db    794 CTGTAACAGATTCTGCATGGAGAGCTGCATGTACCAACACAAGCTGCGGATCTGCGTCTG 853

Qy    857 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaatttcttacc 916
Db    854 TATCTTCCTACGCCAATACAGCAGGAAATGTTTACTACAGAAGTCCGTCAGGTTCATTAC 913

Qy    917 tgtatgacaacaatctgatcaataccaactgtgtactgactaaattttcactgctttccg 976
Db    914 TGTATGATAACAACTATGTGAATACCAATTGTGTATTAAACATATTCTCATCCCTTTCAG 973

Qy    977 gatgttctccttcacctgcaccggatgtatccagctgtggattttaatt 1025
Db    974 GATGTTCTCCTTCCCCAGCACCAAGTGTAGCAAGCTGTGGATTTTAATT 1022


RESULT   6
AR159957
LOCUS       AR159957                 555 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 5 from patent US 6251651.
ACCESSION   AR159957
VERSION     AR159957.1  GI:16222831
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Yamaguchi,S. and Matsuura,A.
  TITLE     Protein-deamidating enzyme, gene encoding the same, production
            process therefor, and use thereof
  JOURNAL   Patent: US 6251651-A 5 26-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="unknown"
BASE COUNT      159 a    122 c    112 g    162 t
ORIGIN


  Query Match           31.9%;  Score 344.2;  DB 6;  Length 555;
  Best Local Similarity   76.7%;  Pred. No. 3.2e-68;
  Matches  421;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy    472 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 531
Db      7 AGTGTTATTCCTGATCTGGCAACGCTGAACAGTTTATTTACCCAGATCAAAAACCAGGCT 66

Qy    532 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 591
Db     67 TGCGGAACTTCTACAGCATCTTCTCCTTGTATCACCTTCAGATATCCGGTTGACGGATGT 126

Qy    592 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 651
Db    127 TATGCAAGGGCTCACAAAATGAGACAAATCCTATTGAACGCCGGCTATGACTGTGAAAAG 186

Qy    652 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 711
Db    187 CAGTTCGTATATGGTAATCTGAGAGCTTCTACAGGAACATGCTGTGTATCATGGGTATAT 246

Qy    712 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 771
Db    247 CACGTAGCAATTTTGGTAAGCTTCAAAAATGCTTCAGGAATTGTTGAAAAAAGAATCATA 306

Qy    772 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 831
Db    307 GATCCTTCATTATTCTCCAGCGGTCCTGTAACAGATTCTGCATGGAGAGCTGCATGTACC 366

Qy    832 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 891
Db    367 AACACAAGCTGCGGATCTGCGTCTGTATCTTCCTACGCCAATACAGCAGGAAATGTTTAC 426

Qy    892 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 951
Db    427 TACAGAAGTCCGTCAGGTTCATTACTGTATGATAACAACTATGTGAATACCAATTGTGTA 486

Qy    952 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 1011
Db    487 TTAAACATATTCTCATCCCTTTCAGGATGTTCTCCTTCCCCAGCACCAAGTGTAGCAAGC 546

Qy   1012 tgtggattt 1020
Db    547 TGTGGATTT 555


RESULT   7
E59337
LOCUS       E59337                   555 bp    DNA     linear   PAT 07-FEB-2001
DEFINITION  Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof.
ACCESSION   E59337
VERSION     E59337.1  GI:13023304
KEYWORDS    JP 2000050887-A/3.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Shotaro,Y.A.M.M.
  TITLE     Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof
```

```
  JOURNAL   Patent: JP 2000050887-A 3 22-FEB-2000;
            AMANO PHARMACEUT CO LTD
COMMENT     OS   Chryseobacterium gleum
            PN   JP 2000050887-A/3
            PD   22-FEB-2000
            PF   04-JUN-1999 JP 1999158703
            PR
            PI   SHOTARO YAMAGUCHI,AKIRA MATSUURA
            PC   C12N15/09,C12N9/80//A21D2/36,A23C9/154,A23C11/06,A23J3/00, PC
             A23J3/10,
            PC   A23J3/16,A23J3/18,A23J3/34,A23L1/176,A23L1/23,A23L1/317, PC
            (C12N9/80,C12R1:01),(C12N9/80,C12R1:20),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   source          1. .555
            FT                   /organism='Chryseobacterium gleum'.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      159 a    122 c    112 g    162 t
ORIGIN


  Query Match             31.9%;  Score 344.2;  DB 6;  Length 555;
  Best Local Similarity   76.7%;  Pred. No. 3.2e-68;
  Matches  421;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy    472 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 531
          ||||| |||||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db      7 AGTGTTATTCCTGATCTGGCAACGCTGAACAGTTTATTTACCCAGATCAAAAACCAGGCT 66

Qy    532 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 591
          ||||| || ||||| || || || || || ||||| ||||||||||| || |||||||||
Db     67 TGCGGAACTTCTACAGCATCTTCTCCTTGTATCACCTTCAGATATCCGGTTGACGGATGT 126

Qy    592 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 651
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    127 TATGCAAGGGCTCACAAAATGAGACAAATCCTATTGAACGCCGGCTATGACTGTGAAAAG 186

Qy    652 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 711
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    187 CAGTTCGTATATGGTAATCTGAGAGCTTCTACAGGAACATGCTGTGTATCATGGGTATAT 246

Qy    712 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 771
          ||||| ||||| |||||||||| |||||||||| ||| |    |||||||||| ||
Db    247 CACGTAGCAATTTTGGTAAGCTTCAAAAATGCTTCAGGAATTGTTGAAAAAAGAATCATA 306

Qy    772 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 831
          ||||||||| |||| || |||||||||||||||||| | |||||||||   || ||
Db    307 GATCCTTCATTATTCTCCAGCGGTCCTGTAACAGATTCTGCATGGAGAGCTGCATGTACC 366

Qy    832 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 891
          |||||    |||||||||||| || || || || || || ||||| ||||||||||||||
Db    367 AACACAAGCTGCGGATCTGCGTCTGTATCTTCCTACGCCAATACAGCAGGAAATGTTTAC 426

Qy    892 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 951
          ||||||||||      ||| |  |||||||| |||||     | ||||||| ||||||
Db    427 TACAGAAGTCCGTCAGGTTCATTACTGTATGATAACAACTATGTGAATACCAATTGTGTA 486

Qy    952 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 1011
          | |  | ||| |||   ||||| |||||||||||||| || |||||   |||| | |||
Db    487 TTAAACATATTCTCATCCCTTTCAGGATGTTCTCCTTCCCCAGCACCAAGTGTAGCAAGC 546

Qy   1012 tgtggattt 1020
          |||||||||
Db    547 TGTGGATTT 555


RESULT   8
I66494/c
LOCUS       I66494                  7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             5.6%;  Score 60;  DB 6;  Length 7218;
  Best Local Similarity   1.6%;  Pred. No. 0.0014;
  Matches    6;  Conservative  231;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    126 ttccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagct 185
          ||   ||:::::  ::::::  :: ::::::::::::  ::: : ::: ::    :::
Db   1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy    186 gaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggt 245
          ::::::    :::::::  : :  :: ::::: ::: :::::::: :::: :: :::::
Db   1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy    246 gtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagta 305
          :  :   : :  ::  : : ::    : ::::  :::  ::  :::::::: ::: :: :
Db   1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy    306 tatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcc 365
          :  :::: :   ::: :::  :  ::::: :::      : : : :       ::::
Db   1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy    366 taatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaa 425
          :: :: :: :::: :::::::: ::::   : :::   :::::: : :::: :    ::
Db   1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy    426 aacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctga 485
          :: ::    :: :::::::: :::::::: :::  :: :::  :: ::: : ::    ::
Db   1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1082

Qy    486 tgtagctacattaaattc 503
          : ::  : :  ::: ||
Db   1081 RRRRRRRRRRRRRRRATC 1064


RESULT   9
AX346947/c
LOCUS       AX346947                7319 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 2018 from Patent WO0200928.
ACCESSION   AX346947
VERSION     AX346947.1  GI:18494833
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 2018 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .7319
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     1946 a    263 c   1552 g   3558 t
ORIGIN
```

```
  Query Match             4.2%;  Score 45.4;  DB 6;  Length 7319;
  Best Local Similarity   46.7%;  Pred. No. 2.9;
  Matches  179;  Conservative    0;  Mismatches  201;  Indels    3;  Gaps    1;

Qy       5 aaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatga 64
           ||| |||  ||   | || | | |||  ||    || |  |  |||||||||| | ||| |
Db    6212 AAACTAATAAAATTAATTCAAACAAATCCAAATATATATCAATAATTACAATAAATATAA 6153

Qy      65 aaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccg 124
           | ||  |    ||| |||| |     | |   ||| | |||| |   ||  |     ||
Db    6152 ATAAAATAAACTTACCAATTAAAAATTATACTACCATATTAAATAAAAAAACAAAATCCA 6093

Qy     125 attccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagc 184
           |   | |||  ||      ||   ||  ||   |||||  ||    |||| |    ||
Db    6092 A---CTACGTAAAATTTATAAAACACAAAATAAAAAAAACAAAAAAAAATTAAAAAAAAA 6036

Qy     185 tgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaagg 244
             ||| |     ||| |   |     |    | |  || |||||  ||  || |
Db    6035 ATAAAAAACGAAAAAATATATAAAATAAACACCAATAAAAAAAAAACTAAAATATTACCT 5976

Qy     245 tgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagt 304
           |   |||||    || |    ||    |  | | ||     |   || ||||  |  | |
Db    5975 TAATATTTAACAAAAATTTTAAAACAAAAAACAATATTAATAATCAAAAAAACAACTACT 5916

Qy     305 atatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagc 364
           | ||   ||    || |         || ||  ||| |         | ||||      |
Db    5915 AAATAATAAAAAATAAATTCATCAAAAAAAAAAAATATAACAATTTAAATTTCCATTTTC 5856

Qy     365 ctaatagcaatgaaataggaaaa 387
           |||||| |    |||||  ||||
Db    5855 CTAATAACCTCAAAATATAAAAA 5833


RESULT  10
AB018378
LOCUS       AB018378               5993 bp    DNA     linear   PLN 15-OCT-1998
DEFINITION  Glycine max GmENOD93 gene for early nodulin, complete cds.
ACCESSION   AB018378
VERSION     AB018378.1  GI:3763850
KEYWORDS    early nodulin.
SOURCE      Glycine max (cultivar:Akisengoku) root nodules DNA.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 5993)
  AUTHORS   Kouchi,H.
  TITLE     GmENOD40 gene for soybean early nodulin
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 5993)
  AUTHORS   Kouchi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-1998) Hiroshi Kouchi, National Institute of
            Agrobiological Resources, Department of Plant Physiology; 2-1-2,
            Kannondai, Tsukuba, Ibaraki 305, Japan
            (E-mail:kouchih@abr.affrc.go.jp, Tel:81-0298-38-8377)
FEATURES             Location/Qualifiers
     source          1. .5993
                     /organism="Glycine max"
                     /cultivar="Akisengoku"
                     /db_xref="taxon:3847"
                     /tissue_type="root nodules"
     gene            join(2379. .2463,3166. .3227,3598. .3673,3927. .4021)
                     /gene="GmENOD93"
     CDS             join(2379. .2463,3166. .3227,3598. .3673,3927. .4021)
                     /gene="GmENOD93"
                     /codon_start=1
                     /product="early nodulin"
                     /protein_id="BAA33816.1"
                     /db_xref="GI:3763851"
                     /translation="MAKGNSPLERPSLASLDQKLAFAKRCSHEGVLAGAKAAVVASVA
                     SAIPTLASVRMLPWARANLNHTAQALIISTATAAAYFIVADKTVLATARKNSFNQPSN
                     SEA"
     exon            <2379. .2463
                     /gene="GmENOD93"
     intron          2464. .3165
                     /gene="GmENOD93"
     exon            3166. .3227
                     /gene="GmENOD93"
     intron          3228. .3597
                     /gene="GmENOD93"
     exon            3598. .3673
                     /gene="GmENOD93"
     intron          3694. .3926
                     /gene="GmENOD93"
     exon            3927. .>4021
                     /gene="GmENOD93"
BASE COUNT     2229 a    944 c    741 g    2079 t
ORIGIN


  Query Match             4.1%;  Score 44.4;  DB 8;  Length 5993;
  Best Local Similarity   55.9%;  Pred. No. 5;
  Matches  105;  Conservative    0;  Mismatches   81;  Indels    2;  Gaps    1;

Qy       4 taaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatg 63
           ||| || |||||   || |||||   |   |  |  ||| |  |||  | ||  |
Db    1991 TAAGATGACCAAAGTACCTAACAGCCAGGTATAAAGAAAATTAAAAAAAAAAGAAGATAA 2050

Qy      64 aaaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgcc 123
           |||   | |  ||| ||||| ||   ||   || |    ||||  ||  |  |||  ||
Db    2051 AAACTACCTCCTTTTTCAATTATCATCTAGCTGTCTTATTTAA--TTACAACCCTTAGCT 2108

Qy     124 gattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaag 183
           ||| ||| || | |  |||||    |   ||| |  |||| |  |||| ||| ||| |
Db    2109 TATTACAAGGGTATTTTGGAAACATGCCATAAGAATCAACTGACCGTAACTGAATCTCAA 2168

Qy     184 ctgaaaga 191
           ||||| |
Db    2169 GTGAAACA 2176


RESULT  11
STATRSC
LOCUS       STATRSC                13776 bp    DNA     linear   BCT 14-MAR-2000
DEFINITION  Staphylococcus aureus conjugative transfer gene complex (trs).
ACCESSION   L11998
VERSION     L11998.1  GI:310606
KEYWORDS    trsA gene; trsB gene; trsC gene; trsD gene; trsE gene; trsF gene;
            trsG gene; trsH gene; trsI gene; trsJ gene; trsK gene; trsL gene;
            trsM gene; trsN gene.
SOURCE      Staphylococcus aureus.
  ORGANISM  Staphylococcus aureus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 13776)
  AUTHORS   Morton,T.M., Eaton,D.M., Johnston,J.L. and Archer,G.L.
  TITLE     DNA sequence and units of transcription of the conjugative transfer
            gene complex (trs) of Staphylococcus aureus plasmid pGO1
  JOURNAL   J. Bacteriol. 175 (14), 4436-4447 (1993)
  MEDLINE   93322322
FEATURES             Location/Qualifiers
     source          1. .13776
                     /organism="Staphylococcus aureus"
                     /isolate="pG01"
                     /db_xref="taxon:1280"
     gene            complement(161. .347)
                     /gene="trsN"
     CDS             complement(161. .349)
                     /gene="trsN"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA71951.2"
                     /db_xref="GI:7239722"
```

```
                /translation="MNNNEENSVFFGKKKKVSLHLLVDPDMKDEIIKYAQEKDFDNVS
                QAGREILKKGLEQIKSNK"
     gene       521. .1495
                /gene="trsA"
     CDS        521. .1495
                /gene="trsA"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71952.1"
                /db_xref="GI:310608"
                /translation="MDNITFRKRIEEYRKRSIEQKKYESKLNYQSRFGVKDDEVSTFN
                IKGIATDKKENKLVMAEDREDKQKVLLYANKDNDMYYLNSGSTYKEAVQNTFGNETNY
                EMVDDELFIYNSKNQALTNYDQLDEYHNRQKQLQYNTKLEQERKQRLTDTNKVSENAK
                EDLALSLGLTATALKGNINISNVEKVTDNESMVTIENEINGKSEHLISDFKDSFIEKA
                KEKLTVSPSMRLGFEEFKNASGQIAMGFTEFFGARYDESTLENDPKVVATEKGYSVSV
                FDKDGNKKDLTSADTFEEAVKLSKEMKHDNSLSRSEIELNQELEKKKI"
     gene       1512. .1829
                /gene="trsB"
     CDS        1512. .1829
                /gene="trsB"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71953.1"
                /db_xref="GI:310609"
                /translation="MIKKFSLTTVYVAFLSIVLSNITLGAENPGPKIEQGLQQVQTFL
                TGLIVAVGICAGVWIVLKKLPGIDDPMVKNEMFRGVGMVLAGVAVGAALVWLVPWVYN
                LFQ"
     gene       1889. .2296
                /gene="trsC"
     CDS        1889. .2296
                /gene="trsC"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71954.1"
                /db_xref="GI:310610"
                /translation="MSVEQEVKDIMLDMLGEEKKVKENIIPRNVDASYKIFLNLNGKE
                LAIVFVPFLLFLTIAFGVTFLMGLLNLITGFMVFIVGLIFGLTIYGLLTIRPISTKEN
                IRMIDTIKQSQRFSRRQKVYFYKSKEGLGDDDV"
     gene       2283. .2966
                /gene="trsD"
     CDS        2283. .2966
                /gene="trsD"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71955.1"
                /db_xref="GI:310611"
                /translation="MMMFKKKNKQNKNNEQQFEIINRYIKGDGRNQSIQEDSLIAGFY
                KDYVVTKDGYISGGLNITGVNTELLTKFEQNDLLEEYSSFISNILYDEPMIKTITEPV
                NIQNYIKHLYRQVLLDMKNNNNKPTRRGQLIAGRILFLQEKLAKGEMSTKKHYIFFSE
                KITENTVAQLDHAVNRLHEKINLNKEALQQAFSHYGLETRPIYQTEYLNQFYALYDYK
                GYNDYHLSN"
     gene       2981. .4999
                /gene="trsE"
     CDS        2981. .4999
                /gene="trsE"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71956.1"
                /db_xref="GI:310612"
                /translation="MIMAFLKKKKQEQVTNKIHNESFQKLTEVLDTDSVDSIYPFSWI
                EKKSHIETGENYIKNLLVVDYPQSVKGAYLSNLLKKNGNIQITKFIRPANIERMIDHL
                NNSIKNKTAEQMRTTDPKRNATIKREIESSKKQLDKFLDEKTGFMYMYMYITLNGDSY
                EKIQALEKDVKRTLTRLRLKTHTPTNAMRESFHTVLPLNRNFLSAFTQQNMDTATAGH
                FFMFDDSEIIDLTPNTSVFGINKNTDSLVAVDFNNKEKTLNKNMTIIGTSGVGKSTLN
                MRMILDNIKKSIRQFIIDPEDEFSYITKYYGGTVVNISTSSNIKINPFEIFSEEVFEK
                EDETDNETTSDVLTENNEHESQIDTLVRSKIGKLKTFFRVLKDEISQTEISVLSSTLR
                QLYQDKGFKGNAKLSDFKSEDYPTLTELYNKLKDLDPEKYEVLKDLTLIIEDYTMEHG
                TTTIFDGYTNIKLDNEIVTFNLKPLQTEKDVQSAAYLNIFSFLWDEITKDRTTETVLM
                TDELHFLATNEYSLDFYYQAYKRIRKYGGGAIASTQQIKDILRTSQEIGSAIIENSHT
                KFFFGMDNVGVDDVVDKLGLKFSDQEISHLTKKKKGEALLLYGTQRAFIRIDLDREET
                RLWNKELYETIYEEPADVEPNYVEQLGLTDIDLAELEEELRQAEMIYE"
     gene       5011. .6291
                /gene="trsF"
     CDS        5011. .6291
                /gene="trsF"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71957.1"
                /db_xref="GI:310613"
                /translation="MKIRNFNEGTLVGKHNHLEYIINVDLCHISDIDEIRYDLTEPNE
                LFFEVKDLQKEGDYFHIIYNTDNEYNSFLSAKKESKALKLSLMDYVLKVDPLKDNITV
                MHPANLFFKNIEDIKIGFKGHEYLPKPKINNLEQYKLLIMSTLSQYTFDKYYRNKFEV
                LSKEKDDFFHKVNNAETFEKLKEIVRHELNQVQTDYLLNEEKRKRETRKKYIRNLILG
                FVGIAIVISLISFMIINGKQQELDSKIAQADKQEQRSKVYENLYNGNVDQAVKGMKRD
                DSFNKKDIEKTLKKEKKYEELIELSSKNTPYVIEQLYKEGKQNKIRELAFNFENNDTL
                SLEKKILDKDGTAFSVGSTGKYEEQSKRLALASAKEGYVESAKDINKKLKDNEVEEEI
                NKAEIKQLKEEKDSTKDKDKKKEIQKDIEELEKK"
     gene       6309. .7385
                /gene="trsG"
     CDS        6309. .7385
                /gene="trsG"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71958.1"
                /db_xref="GI:310614"
                /translation="MWSAIAKGITMLMKKKALSKILKPVIAIVIGFFLLIFIMIAYLV
                GSFTKEINESGEQKMKEAESTATGACTGGSVDGSGISAFEKNAKGGALEGKGKDMVKI
                AKKNKIPPKLFIAIVASESEWGRGANATKQKNPLSIMGAGPLQVYPSIEEGLDKGAKN
                LYDLYISEGLTTPEKIGPKYAPVGASNDPDDLNSNWVPTVKKIMKSFGGKEAKCSTES
                SGGGSDGKGFDFKGEFPKPDKSKYNGQSYPWGQCTWYVHQRRKEIGKPVPLTWGNGGD
                WGDNAKAQGWEVGSKPKAGRGASVKPGNFGAPPPYGHIMFVEKVKKDGGIVVSEANVK
                GLGVISSREFSKAETQRMQFIYDK"
     gene       7395. .7880
                /gene="trsH"
     CDS        7395. .7880
                /gene="trsH"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71959.1"
                /db_xref="GI:310615"
                /translation="MKKLLTIVILFTLVLAGCGQNQSEKNSEKHNEEKTVDTTEERND
                MIKGLIEYGYENDDYKGYKHMKDILSEKLQKQAELQNQSAQTKGVKKETKDIKLYQDV
                DNDDELIYSIKLSITNKKEKNSDYSMRYGKIELTEEEGKQKIDNVEELTQRNLVSMNQ
                E"
     gene       7896. .9998
                /gene="trsI"
     CDS        7896. .9998
                /gene="trsI"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA71960.1"
                /db_xref="GI:310616"
                /translation="MNTLILCEKPSQAMDLSTVFAKKKKQNGYMEISDEQLNVSGFLT
                WAVGHLVELKEPQEYDEKYKNFSTYPILLEKDDFQFKVSDKTKDQFNNIKKIFKENKI
                DEVIIATDPAREGENIAYKILNQLKVTDKVTIKRLWLTSKVESSIRKAFKNILPKEKT
                YGFYKEGRARELSDWLVGINLSRHFTKISRELGNDGVIHIGRVSSPTLNMVYNRENNI
                KGFKGKKFYKVSATINKDEQEVKTELKNKFDSEDELHEFLFENDITDLTQKGLVTDIE
                KEIGYTMPPKFYDLSALQEDMNDKYKISAKRTLEIAQTLYEKKLITYPRTDSRYITED
                EKEMLLENIDYLKEITKINLNNELTNNSLINPSKIEDHYAILITGNDFNKVDLKEEEI
                NVYKSILQNVAMNFMDKEQYETTTIEIAVKKLMFEVKGKIIQDNGFKALLNKQKTSEE
                TIPNFEKNEEVDIELDLLEKETTPPKRYTEKTLLKAMANPIETLEDEGLKSTLKEVKG
                LGTPATRADIIENLKKNKYIQVQKNKIYITKNGILACLLLEGHLLSKPDLTGQWEKYL
                NGISKGEKDDDSFINTINEMIKKTINEEVKNKESIQKVAKEKVSTNNIAKCPACDNGY
                LIDRKGFYGCTEYNNGCEFTIPKKLLEKSIPPTVVKALCESKTTKKLKGFKSKKSGKS
                FDCKLILTKENKLQFSFD"
     gene       10014. .10478
                /gene="trsJ"
```

```
     CDS             10014. .10478
                     /gene="trsJ"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA71961.1"
                     /db_xref="GI:310617"
                     /translation="MKKGIFILTGIIWLVLIYLAQILAYALNTLSNTSNKQKIFSDIS
                     QLFIPPAKHPISTIIGFVSDKNTVFFVIGVIATIICIYILIKQLFESKDGNEKNQDYK
                     IAKHGSHGSAKFAEDSELFSEGHYKKMKENDITNYVYNSLDKSLIEERSDKE"
     gene            10475. .12115
                     /gene="trsK"
     CDS             10475. .12115
                     /gene="trsK"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA71962.1"
                     /db_xref="GI:310618"
                     /translation="MTTLLGELESKATSINKKNLVIQDFDTKCPNRNIFVVGGPGSYK
                     SAGYVIPNVIVKNQQSIVVTDPSGEVYEKTSNIKRMQGFDVRVVNFKNFLASDRQNFF
                     DYIDKDSDCSIVAELIIKSAGDSKINKDVWYKASVGLLNSLLLYAKYEFEPEKRTIGN
                     IIKFIQNNKPNQDEEGSVELDKRFNELSKDHPARESYEFGFAVSEGRTRASIILTFVA

  Query Match             4.1%;  Score 44.4;  DB 1;  Length 13776;
  Best Local Similarity   46.7%;  Pred. No. 4.7;
  Matches  141;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     224 aagaaaacggaatgataaaggtgtcatttatgttaactgcgcaattctatgaaattaagc 283
           ||||||    || ||| ||  ||     |   ||||||    |   ||     | |
Db    8935 AAGAAATAACAAAGATTAATTTGAATAATGAATTAACTAATAACAGCTTGATTAATCCTT 8994

Qy     284 cgaccaaagaaaatgagcagtatatcggaatgcttagacaggctgttaagaatgaatctc 343
           | |  | |||| || |  |   ||   |||   | || | | | |||| || | |  |
Db    8995 CTAAAATAGAAGATCATTACGCAATATTAATAACTGGAAATGATTTTAACAAAGTAGATT 9054

Qy     344 ctgtacacattttcttaaagcctaatagcaatgaaataggaaaagtggagtctgcaagtc 403
           | |        | ||   | |||  | |  | ||  ||| ||||          |
Db    9055 TAAAAGAAGAAGAAATCAATGTTTATAAAAGTATATTACAAAATGTGGCCATGAATTTTA 9114

Qy     404 cggaagacgtaagatattttaaaacgatcctgacaaaagaagtaaaagggcaaaccaata 463
           |||  | | |  |||   |  || |   |   ||  |  || |||  ||  |    |
Db    9115 TGGATAAAGAACAATATGAAACGACTACTATAGAAATCGCTGTTAAAAAGCTTATGTTTG 9174

Qy     464 aattggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaaga 523
           || ||    ||  ||||  | | |       ||| ||  |||| || ||| ||  ||| |
Db    9175 AAGTGAAAGGTAAAATTATTCAAGATAACGGATTTAAAGCTTTGTTAAATAAACAAAAAA 9234

Qy     524 at 525
           |
Db    9235 CT 9236


RESULT  12
AF051917
LOCUS       AF051917               46445 bp    DNA     circular BCT 30-SEP-1998
DEFINITION  Staphylococcus aureus plasmid pSK41, complete sequence.
ACCESSION   AF051917 L19570
VERSION     AF051917.1  GI:3676412
KEYWORDS    .
SOURCE      Staphylococcus aureus.
  ORGANISM  Staphylococcus aureus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 23598 to 41563)
  AUTHORS   Firth,N., Ridgway,K.P., Byrne,M.E., Fink,P.D., Johnson,L.,
            Paulsen,I.T. and Skurray,R.A.
  TITLE     Analysis of a transfer region from the staphylococcal conjugative
            plasmid pSK41
  JOURNAL   Gene 136 (1-2), 13-25 (1993)
  MEDLINE   94123990
REFERENCE   2  (bases 1 to 46445)
  AUTHORS   Berg,T., Firth,N., Apisiridej,S., Hettiaratchi,A., Leelaporn,A. and
            Skurray,R.A.
  TITLE     Complete nucleotide sequence of pSK41: evolution of staphylococcal
            conjugative multiresistance plasmids
  JOURNAL   J. Bacteriol. 180 (17), 4350-4359 (1998)
  MEDLINE   98389645
REFERENCE   3  (bases 23598 to 41563)
  AUTHORS   Berg,T., Firth,N., Apisiridej,S., Hettiaratchi,A. and Skurray,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-1993) School of Biological Sciences, University
            of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
REFERENCE   4  (bases 1 to 46445)
  AUTHORS   Berg,T., Firth,N., Apisiridej,S., Hettiaratchi,A. and Skurray,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAR-1998) School of Biological Sciences, University
            of Sydney, Sydney, NSW 2006, Australia
COMMENT     On Oct 1, 1998 this sequence version replaced gi:405558.
FEATURES             Location/Qualifiers
     source          1. .46445
                     /organism="Staphylococcus aureus"
                     /db_xref="taxon:1280"
                     /plasmid="pSK41"
     CDS             72. .686
                     /note="Orf204"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC61931.1"
                     /db_xref="GI:3676413"
                     /translation="MGNLDFDHEGYRKLLPLKDFKHFIKHSDSKTRNHFIGMAFNNKT
                     DVFSRNFSNKMINIFPKNEFYIHIQQFPENDSPSYMVYNTDITAFIQFRLVGRSENKY
                     LENIIRDYGVLIEVEALHSFKKGMGAMLVSQLEKLSDKLFIPIYLYDTNLKDELYYQD
                     LGFFDTTKKGNHGEPLLVYKPKNLETNKKEKPIKSLFKKVFNIN"
     CDS             755. .2026
                     /note="Orf423"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC61932.1"
                     /db_xref="GI:3676414"
                     /translation="MDYTNQKRKYIIFGIVLAITLLIVGIAYYLISLNNAKGQVKEFE
                     KLVQHNDYNSIAKKLSNKEETLSENDAKNFVLFVKSKENQEQYNEQIQKIKSNIDKNK
                     DNNTVYGSITDQYGNNLITIKKNGKTFFFIDKLEFKPKLTSVYVKEFNNTGIYKYQQE
                     NAKKVIAESNQTTELGKFFIGKYSIKTDKTIEDSIHNGTLTGKIDIDLSNRGNNNKIY
                     ADEDFNQSWFKVKLNNDSLLEKNSYKLLIDDKDVDYNSKKTYGKYYNPTELSVYAIGK
                     LEGKEFKTNRINIRRNYDNKPQNLKLSFKESQIRDYESKSKKVKEKAKSYIKEYTKEL
                     NKAYKKKDYKYISDYVKNNSQLEKQLKKQVKSKSKVKFKDVKINNVKRDKNKIYIELS
                     KTNNGNVFQTKYTLDYDNVSNEFLVEDSHDN"
     sig_peptide     755. .862
                     /evidence=not_experimental
     mat_peptide     863. .2023
                     /product="unknown"
     CDS             complement(2068. .2484)
                     /note="Orf149"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC61933.1"
                     /db_xref="GI:3676415"
                     /translation="MKIINPNMPEPYKYETDYRNIPREYLNPHIPKGRGIVKWNAFKT
                     IPQQYEILEQHIEDQHKINMPLLSDDQLQEINEKVSSKISKNILTDLDYWKDGYIYSI
                     QCYINKIDEFNGFMIITNSNSNNKLKIPLSSIVSVQ"
     gene            2779. .3336
                     /gene="res"
     CDS             2779. .3336
                     /gene="res"
                     /codon_start=1
                     /transl_table=11
                     /product="putative resolvase Res"
                     /protein_id="AAC61934.1"
                     /db_xref="GI:3676416"
                     /translation="MANIGYARVSTRDQSLDLQIDALSNFGCEKIFSEKVSGRKVNRT
                     ELDKCLDYLREGDTLVIYKLDRLGRTTKQLIDLAQWLENNNIELQIIDMNINTKDPMG
```

```
                KMFFTMMSAFAELEANLLSERTKKGLESARARGNMGGRKAMPQEKKEYIKHLYDTKEY
                TGQEIAKKTNVSRDTVYRILREFKK"
     CDS        3618. .4397
                /note="Orf259"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC61935.1"
                /db_xref="GI:3676417"
                /translation="MEVVRLNQNLFNKLRGNEISSNKNGSRPYYYSFKRNNNRVCIPF
                RTNAQKVPNKYKINLGGEQPDKPNSAIDLTKSIVISNDEYLNNRSKAKIPQNVNNFLK
                QQAPAIEQKYDTMSNDYIKAKASLSKIPLVKYSTMQYFHKELNIQDSIDNQQTKNAIN
                ELISNGKSNKYNKLQSSLPNEKLNLLDDYETLYEFKSLTDYPAKINSNDIDNPFLEVE
                KNNKHFTLSALTIKNEPEKHVKDFLNYDIENEKNKDIDLDL"
     CDS        4531. .6147
                /note="Orf538"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC61936.1"
                /db_xref="GI:3676418"
                /translation="MRFTPRGILNNTDYELNSEFYYVPASDKQMNDQRDMTNKSIVGN
                LVYRNNFYNEVYTQPVYDLDDSNEKYQTPHIISIDDCEFDYNLTNLDDYFTKDTANKG
                YEKLVNAIKSEEITYFDSDINEVDDFTQQDEYSNDWYTNTFVDLVLSMKGYSYHVQPN
                RESVIEDSHIITENDITYNVSYNPEKHHLYINDFPAIPIEENDNIRLLNDNLTQDMEK
                TYDNINYNFSYNDFKTLEDDSYIYIDLTDYKDNEGKFVKNLIEDIPPNEIIKDRLKQN
                INLFDERANYNALAVYRRDLEDFSDRNNLDSLDKVKDLKTFYDTEINIKGLNINNDIE
                KKALNNQLAFNLYNEPLNNMFRDNNVLYEVYEKEIYDEKIPLSIEDRKEVKTYLDDAR
                NNFSNNFYEYTTNKSSILAENSLTVLDKAKYDLEFAPVSNHLYKERENFYEGMKQATL
                PLHNITKEQYKERENDIEIIGHFKNYINTELHGTKSSEFKPNFVNDSLENETNSIAFE
                DIINGSEETHSNIINSIKNELKQERDNLEM"
     CDS        6350. .8074
                /note="Orf575; similar to Lactococcus lactis pRS01 LtrC"
                /codon_start=1
                /transl_table=11
                /product="LtrC-like protein"
                /protein_id="AAC61937.1"
                /db_xref="GI:3676419"
                /translation="MAYSKGNYKKKKIDSIIEELNNKLDDFKNNDKTYKEFLDTTAKF
                HNYSINNIILIAGQNPEATAVAGYKAWKNKFDRQVQKGAKSMNIIAPIVQKRNVEIKD
                DNGNIVRDNNGNPKTERKPVITGYRAHNVFDVADTKGKPLITAKDLIKTEFENSNDYK
                DLYNEFKDYINEELTPSVEEKHFLDDPTLSNGAKGYYSPKSDEIVISDDLSYDMRFKT
                LIHEYAHSQLHNNDIGKTQISEHSRSLKEIEAESSAYVVANYYGLDTSDYSLGYLSGW
                GHNISDDELKAHIKNIHSFAKTTIEEINSLPEFSQYIDKKLESEMNKDVYKDLSTMID
                TNLKNGFDKITIIKGNLVNDYGLNEISENSYENDDFKVNIDYKGFNTNNSQDQAKIEL
                INKHDDSLNRDYNFTQTYNRNVINNTTTIFVQDDDNEDKNKKYIHERDIDGNILEEKH
                KLSPQDEIVSFENFVNKSVEENGLKDTLRQFVSNGVLMGYELNVDNIDNTNETYFSMN
                KNQQNGQRSNMTAKLEHDTNENVYIDFKIKNSAGIKNISYTESNEEFNKDEQNKEIEN
                KQEEEIDI"
     gene       complement(8115. .10112)
                /gene="nes"
     CDS        complement(8115. .10112)
                /gene="nes"
                /codon_start=1
                /transl_table=11
                /product="oriT nickase Nes"
                /protein_id="AAC61938.1"
                /db_xref="GI:3676420"
                /translation="MAMYHFQNKFVSKANGQSATAKSAYNSASRIKDFKENEFKDYSN
                KQCDYSEILLPNNADDKFKDREYLWNKVHDVENRKNSQVAREIIIGLPNEFDPNSNIE
                LAKEFAESLSNEGMIVDLNIHKINEENPHAHLLCTLRGLDKNNEFEPKRKGNDYIRDW
                NTKEKHNEWRKRWENVQNKHLEKNGFSVRVSADSYKNQNIDLEPTKKEGWKARKFEDE
                TGKKSSISKHNESIKKRNQQKIDKMFNEVDDMKSHKLNAFSYMNKSDSTTLKNMAKDL
                KIYVTPINMYKENERLYDLKQKTSLITDDEDRLNKIEDIEDRQKKLESINEVFEKQAG
                IFFDKNYPDQSLNYSDDEKIFITRTILNDRDVLPANNELEDIVKEKRIKEAQISLNTV
                LGNRDISLESIEKESNFFADKLSNILEKNNLSFDDVLENKHEGMEDSLKIDYYTNKLE
                VFRNAENILEDYYDVQIKELFTDDEDYKAFNEVTDIKEKQQLIDFKTYHGTENTIEML
                ETGNFIPKYSDEDRKYITEQVKLLQEKEFKPNKNQHDKFVFGAIQKKLLSEYDFDYSD
                NNDLKHLYQESNEVGDEISKDNIEEFYEGNDILVDKETYNNYNRKQQAYGLINAGLDS
                MIFNFNEIFRERMPKYINHQYKGKNHSKQRHELKNKRGMHL"
     misc_feature complement(10221. .10222)
                /note="putative origin of transfer nic site"
     CDS        10316. .10588
                /note="Orf90"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC61939.1"
                /db_xref="GI:3676421"
                /translation="MADIQKMLEQKEKLQEKIKEEKKKEYEKFGRWFFNKFKVNRSAD
                AKKIINKNLLNIEDTKEHNLERNEIIEETKNSNEENKEINEINHNI"
     CDS        10692. .10925
                /note="Orf77"
                /codon_start=1
                /transl_table=11
                /product="putative membrane protein"
                /protein_id="AAC61940.1"
                /db_xref="GI:3676422"
                /translation="MILLSTAGENAVHINPIFRFFLYFIMCVPMTVALCTIFLKFISI
                FILPWKVCLYIGIALGIIVTIFLLVITIGIEMT"
     CDS        11188. .12228
                /note="Orf346"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC61941.1"
                /db_xref="GI:3676423"
                /translation="MSNVYVMALDFGNGFVKGKINDEKFVIPSRIGRKTNENNQLKGF
                VDNKLDVSEFIINGNNDEVLLFGNDLDKTTNTGKDTASTNDRYDIKSFKDLVECSIGL
                LAREVPEEVVNVVIATGMPSNEIGTDKQAKFEKLLNKSRLIEIDGIAKTINVKGVKIV
                AQPMGTLLDLNMENGKVFKAFTEGKYSVLDFGSGTTIIDTYQNMKRVEEESFVINKGT
                IDFYKRIASHVSKKSEGASITPRMIEKGLEYKQCKLNQKTVIDFKDEFYKEQDSLIEE
                VMSNFEITVGNINSIDRIIVTGGGANIHFDSLSHYYSDVFEKADDSQFSNVRGYEKLG
                ELLKNKVEQESK"
     CDS        12231. .12560
                /note="Orf109"
                /codon_start=1
                /transl_table=11
                /product="unknown"

  Query Match             4.1%;  Score 44.4;  DB 1;  Length 46445;
  Best Local Similarity   46.7%;  Pred. No. 4.2;
  Matches  141;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     224 aagaaaacggaatgataaaggtgtcatttatgttaactgcgcaattctatgaaattaagc 283
           ||||||    || ||| ||  ||      |   ||||||    |   ||     | |
Db   32538 AAGAAATAACAAAGATTAATTTGAATAATGAATTAACTAATAACAGCTTGATTAATCCTT 32597

Qy     284 cgaccaaagaaaatgagcagtatatcggaatgcttagacaggctgttaagaatgaatctc 343
           | |  | |||| || |  |    ||  |||   | || | | | |||| || | |  |
Db   32598 CTAAAATAGAAGATCATTACGCAATATTAATAACTGGAAATGATTTTAACAAAGTAGATT 32657

Qy     344 ctgtacacatttttcttaaagcctaatagcaatgaaataggaaaagtggagtctgcaagtc 403
             | |        | ||   | |||  | |  | ||  ||| ||||          |
Db   32658 TAAAAGAAGAAGAAATCAATGTTTATAAAAGTATATTACAAAATGTGGCCATGAATTTTA 32717

Qy     404 cggaagacgtaagatattttaaaacgatcctgacaaaagaagtaaaagggcaaaccaata 463
           |||  | | |  ||||   |  || |   |   ||  |  || |||  ||  |    |
Db   32718 TGGATAAAGAACAATATGAAACGACTACTATAGAAATCGCTGTTAAAAAGCTTATGTTTG 32777

Qy     464 aattggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaaga 523
           || ||    ||  |||| | | |       ||| ||  |||| || ||| ||  ||| |
Db   32778 AAGTGAAAGGTAAAATTATTCAAGATAACGGATTTAAAGCTTTGTTAAATAAACAAAAAA 32837

Qy     524 at 525
             |
Db   32838 CT 32839


RESULT  13
ATCHRIV40
LOCUS       ATCHRIV40             197405 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 40.
ACCESSION   AL161540
VERSION     AL161540.2  GI:7268224
KEYWORDS    .
```

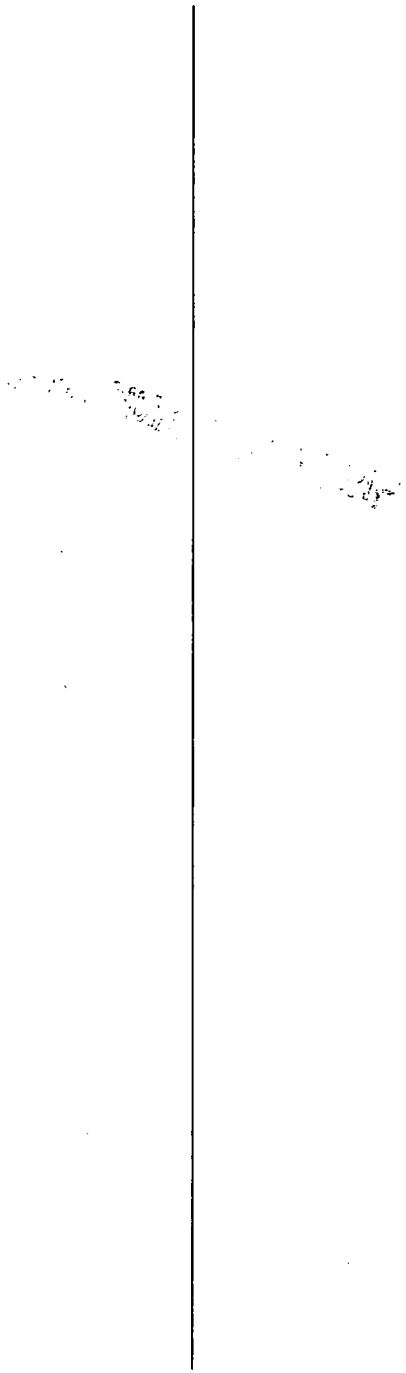
```
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 197405)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV39 at the 5' end and an
            overlap with ATCHRIV41 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .197405
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     exon            complement(8150. .8280)
                     /gene="AT4g14790"
                     /number=1
     gene            8150. .10349
                     /gene="AT4g14790"
     gene            complement(join(8150. .8280,8363. .8531,8621. .8698,
                     8884. .9006,9126. .9246,9299. .9539,9618. .9739,9883. .10162,
                     10286. .10349))
                     /gene="AT4g14790"
     CDS             complement(join(8150. .8280,8363. .8531,8621. .8698,
                     8884. .9006,9126. .9246,9299. .9539,9618. .9739,9883. .10162,
                     10286. .10349))
                     /gene="AT4g14790"
                     /note="similarity to probable RNA helicase SUV3,
                     ATP-dependent, yeast (Saccharomyces cerevisiae)
                     Contains ATP/GTP-binding site motif A (P-loop)
                     AA136-143;EF-hand calcium-binding domain AA383-395"
                     /codon_start=1
                     /product="RNA helicase like protein"
                     /protein_id="CAB78521.1"
                     /db_xref="GI:7268225"
                     /translation="MAYSVVRLRKVSALGISRVLQESLNLYELCSFYICFCVVSRRRL
                     DWFNVVSYDGFVLPNVSADKGSLWRFHFEPEFGDLLRLGVLTRNYRKNSGSPKFDFTG
                     TGTTSKFDFTDLTCPHTWYPIARKKKRKVILHVGPTNSGKTYSALKHLEQSSSGVYCG
                     PLRLLAWEVAKRLNKANVPCDLITGQEKDLVEGATHKAVTVEMADVTSVYDCAIIDEI
                     QASLARLWLKKSTRTFCLGFAFTRALLGIAADELHLCGDPAVVPLVEDILKVTGDDVE
                     VHTYERLSPLVPLKVPVSSVSSIKTGDCLVTFSRKDIYAYKKTIERAGKHLCSVVYGS
                     LPPETRTAQATRFNDETNDFDVLVASDAIGMGLNLNISRIIFSTLQKYDGSETRDLTV
                     SEIKQIAGRAGRFQSKFPIGEVTCLHKEDLPLLHSSLKSPSPILEANLWH"
     intron          complement(8281. .8362)
                     /gene="AT4g14790"
                     /number=1
     exon            complement(8363. .8531)
                     /gene="AT4g14790"
                     /number=2
     intron          complement(8532. .8620)
                     /gene="AT4g14790"
                     /number=2
     exon            complement(8621. .8698)
                     /gene="AT4g14790"
                     /number=3
     intron          complement(8699. .8883)
                     /gene="AT4g14790"
                     /number=3
     exon            complement(8884. .9006)
                     /gene="AT4g14790"
                     /number=4
     intron          complement(9007. .9125)
                     /gene="AT4g14790"
                     /number=4
     exon            complement(9126. .9246)
                     /gene="AT4g14790"
                     /number=5
     intron          complement(9247. .9298)
                     /gene="AT4g14790"
                     /number=5
     exon            complement(9299. .9539)
                     /gene="AT4g14790"
                     /number=6
     intron          complement(9540. .9617)
                     /gene="AT4g14790"
                     /number=6
     exon            complement(9618. .9739)
                     /gene="AT4g14790"
                     /number=7
     intron          complement(9740. .9882)
                     /gene="AT4g14790"
                     /number=7
     exon            complement(9883. .10162)
                     /gene="AT4g14790"
                     /number=8
     intron          complement(10163. .10285)
                     /gene="AT4g14790"
                     /number=8
     exon            complement(10286. .10349)
                     /gene="AT4g14790"
                     /number=9
     gene            10918. .12578
                     /gene="AT4g14800"
     exon            10918. .11075
                     /gene="AT4g14800"
                     /number=1
     CDS             join(10918. .11075,11167. .11290,12261. .12578)
                     /gene="AT4g14800"
                     /note="strong similarity to proteasome chain HsC7-I
                     protein, human
                     contains EST gb:AI993573.1, Y09359, R65485, T22127,
                     F13874"
                     /codon_start=1
                     /product="proteasome chain protein"
                     /protein_id="CAB78522.1"
                     /db_xref="GI:7268226"
                     /translation="MECVFGLVGNGFAIVAADTSAVHSILLHKNKEDKIMTLDSHKLV
                     AASGEPGDRVQFTEYVQKNVSLYQFRNGIPLSTAAAANFTRGELATALRKNPYSVNIL
                     MAGYDKEAGASLYYIDYIATLHKVDKGAFGYGSYFSLSTMDRHYRSDMSVEEAIELVD
                     KCILEIRSRLVIAPPNFVIKIVDKDGAREYGWRISTADA"
     intron          11076. .11166
                     /gene="AT4g14800"
                     /number=1
     exon            11167. .11290
                     /gene="AT4g14800"
                     /number=2
     intron          11291. .12260
                     /gene="AT4g14800"
                     /number=2
     exon            12261. .12578
                     /gene="AT4g14800"
                     /number=3
     exon            12928. .13042
                     /gene="AT4g14810"
                     /number=1
     gene            12928. .15983
                     /gene="AT4g14810"
     CDS             join(12928. .13042,13394. .13501,13559. .13702,15165. .15253,
                     15348. .15365,15846. .15983)
                     /gene="AT4g14810"
                     /note="similarity to Fun12p, Saccharomyces cerevisiae
                     contains EST gb:AA042489, AI998638.1, T88290"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB78523.1"
```

```
                     /db_xref="GI:7268227"
                     /translation="MMIVDGNQKFLILRESLEKYCRTGSGGGGGGGARVAVSDLRKRN
                     QSRSIQRESKDHRLPEKITETVSGARMNGGGDVRKARREHDELLRTSQNRWRRRRCLL
                     RRGYGGGGGEGDREYEDENAVSIYLRTPKKTPTQIKPINSNNHIKMNNSQQSNFNQVL
                     DGGGSQLGINVNETQALALPKACHVETPPASRCHSKINSLSMH"
     repeat_region   13001. .13024
     intron          13043. .13393
                     /gene="AT4g14810"
                     /number=1
     exon            13394. .13501
                     /gene="AT4g14810"
                     /number=2
     intron          13502. .13558
                     /gene="AT4g14810"
                     /number=2
     exon            13559. .13702
                     /gene="AT4g14810"
                     /number=3
     intron          13703. .15164
                     /gene="AT4g14810"
                     /number=3
     exon            15165. .15253
                     /gene="AT4g14810"
                     /number=4
     intron          15254. .15347
                     /gene="AT4g14810"
                     /number=4
     exon            15348. .15365
                     /gene="AT4g14810"
                     /number=5
     intron          15366. .15845
                     /gene="AT4g14810"
                     /number=5
     exon            15846. .15983
                     /gene="AT4g14810"
                     /number=6
     gene            18314. .20558
                     /gene="AT4g14820"
     exon            complement(18314. .19925)
                     /gene="AT4g14820"
                     /number=1
     gene            complement(join(18314. .19925,20002. .20558))
                     /gene="AT4g14820"
     CDS             complement(join(18314. .19925,20002. .20558))
                     /gene="AT4g14820"
                     /note="similarity to 130 KD LEUCINE-RICH PROTEIN (LRP 130)
                     (GP130) L130_HUMAN
                     Contains Cytochrome c family heme-binding site signature
                     AA682-687;Prokaryotic membrane lipoprotein lipid

  Query Match             4.1%;  Score 44.4;  DB 8;  Length 197405;
  Best Local Similarity   49.6%;  Pred. No. 3.7;
  Matches  114;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy     303 gtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaa 362
           || | |   | | |  | | |     ||  | |  |||  |   |    | ||| |   |
Db   82097 GTCTTTGTTAGTTCAAAAAAAAAAGTTTTTGTAATAATAACTCTTTTGAAGTTTGTACGA 82156

Qy     363 gcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatattt 422
           | |||| |  ||  ||| |||||||| | || |  | | | ||  |  ||  ||| ||  |
Db   82157 ACGTAAGAAAAAAAAAAAAGGAAAACTAGAATTAGTATGACCCAATAACTGAAGCTAGGT 82216

Qy     423 taaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcc 482
           | |   |  | ||  | |||||   ||||  ||||  || ||   | | |   |||   |
Db   82217 TTACCAGGCCGTGTGATAAGAATATAAAGCCCAAATTAAAAATGGGTCTAACCTAAAGGC 82276

Qy     483 tgatgtagctacattaaattctttattcaatcaaataaagaatcagtctt 532
             |   |   | | ||  |||| |   ||  ||  |   | || | | ||
Db   82277 CCAATAACTGAAACTAGGTTCTCTCCGCATCCATTTTCTGCATTACTATT 82326


RESULT  14
ATFCA2

LOCUS       ATFCA2                202860 bp    DNA     linear   PLN 29-JUN-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
            No. 2.
ACCESSION   Z97337
VERSION     Z97337.2  GI:5302774
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 202860)
  AUTHORS   Bevan,M., Stiekema,W., Murphy,G., Wambutt,R., Pohl,T., Terryn,N.,
            Kreis,M., Kavanagh,T., Entian,K.D., Rieger,M., James,R.,
            Puigdomenech,P., Hatzopoulos,P., Obermaier,B., Duesterhoft,A.,
            Jones,J., Palme,K., Ansorge,W., Delseny,M., Bancroft,I.,
            Mewes,H.W., Schueller,C. and Chalwatzis,N.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202860)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     On Jun 30, 1999 this sequence version replaced gi:2244829.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATFCA1 at the 5' end and an
            overlap with ATFCA3 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .202860
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     source          1. .70917
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="BAC TAMU2D4"
     gene            13673. .13975
                     /gene="dl3405w"
     exon            13673. .13975
                     /gene="dl3405w"
                     /number=1
     CDS             13673. .13975
                     /gene="dl3405w"
                     /note="weak similarity to N-methyl-D-aspartate receptor
                     glutamate-binding chain, rat"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB10252.1"
                     /db_xref="GI:2244830"
                     /db_xref="SPTREMBL:O23329"
                     /translation="MAKSDIETGGGNELYPGMKESSELRWAFIRKLYSILSLQLLVTV
                     GVSAVVYFVRPIPEFITETHRGLAVFFVILLLPLLRYVSFHILLFIHSFLGSIIVA"
     gene            complement(join(15633. .15841,16271. .16328,16682. .16731,
                     16815. .17057,17098. .17157,17295. .17739,18223. .18333,
                     18701. .19012))
                     /gene="dl3410c"
     exon            complement(15633. .15841)
                     /gene="dl3410c"
                     /number=1
     gene            15633. .19012
                     /gene="dl3410c"
     CDS             complement(join(15633. .15841,16271. .16328,16682. .16731,
                     16815. .17057,17098. .17157,17295. .17739,18223. .18333,
                     18701. .19012))
                     /gene="dl3410c"
                     /note="weak similarity to DNA-binding protein"
```

```
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB10253.1"
                     /db_xref="GI:2244831"
                     /db_xref="SPTREMBL:O23330"
                     /translation="MEKLMVPTWRPDPVYRPPETPLEPMEFLARSWSVSALEVSKALT
                     PPNPQILLSKTEEEEEEEPISSVVDGDGDTEDTGLVTGNPFSFACSETSQMVMDRILS
                     HSQEVSPRTSGRLSHSSGPLNGSLTDSPPVSPPESDDIKQFCRANKNSLNSVNSQFRS
                     TAATPGPITATATQSKTVGRWLKDRREKKKEETRAHNAQIHAAVSVAGVAAAVAAIAA
                     ATAASSSCGKDEQMAKTDMAVASAATLVAAQCVEAAEVMGAEREYLASVVSSAVNVRS
                     AGDIMTLTAGAATDNDSIDLGIGSKLELYEGIRTLRGVQTLKARAMKEVWNIASVIPM
                     DKGLTSTGGSSNNVNGSNGSSSSSHSGELVQQENFLGTCSREWLARGCELLKRTRKGD
                     LHWKIVSVYINKMNQVMLKMKSRHVGGTFTKKKKNIVLDVIKNVPAWPGRHLLEGGDD
                     LRYFGLKTVMRGDVEFEVKSQREYEMWTQGVSRLLVLAAERKFRM"
intron               complement(15842. .16270)
                     /gene="dl3410c"
                     /number=1
exon                 complement(16271. .16328)
                     /gene="dl3410c"
                     /number=2
intron               complement(16329. .16681)
                     /gene="dl3410c"
                     /number=2
exon                 complement(16682. .16731)
                     /gene="dl3410c"
                     /number=3
intron               complement(16732. .16814)
                     /gene="dl3410c"
                     /number=3
exon                 complement(16815. .17057)
                     /gene="dl3410c"
                     /number=4
intron               complement(17058. .17097)
                     /gene="dl3410c"
                     /number=4
exon                 complement(17098. .17157)
                     /gene="dl3410c"
                     /number=5
intron               complement(17158. .17294)
                     /gene="dl3410c"
                     /number=5
exon                 complement(17295. .17739)
                     /gene="dl3410c"
                     /number=6
intron               complement(17740. .18222)
                     /gene="dl3410c"
                     /number=6
exon                 complement(18223. .18333)
                     /gene="dl3410c"
                     /number=7
intron               complement(18334. .18700)
                     /gene="dl3410c"
                     /number=7
exon                 complement(18701. .19012)
                     /gene="dl3410c"
                     /number=8
repeat_region        22713. .22757
                     /note="45bp AT tandem repeat"
repeat_region        34748. .34771
                     /note="24bp CA tandem repeat"
gene                 35515. .36969
                     /gene="dl3415w"
exon                 35515. .35871
                     /gene="dl3415w"
                     /number=1
CDS                  join(35515. .35871,36297. .36857,36943. .36969)
                     /gene="dl3415w"
                     /note="weak similarity to actin-depolymerizing protein
                     N_WASP, brain, bovine"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB10254.1"
                     /db_xref="GI:2244832"
                     /db_xref="SPTREMBL:O23331"
                     /translation="MGKTSKWFRSLLTGKKERTKEHIIQSECVFTSSIPGTPKEKRRW
                     SFRRSSATGPPPPACAITLKDSPPPPPPPPPPPPLQQPFVVEIVDNEDEQIKNVSAEE
                     IEEFAAIKIQACYRSHLENEENIKIVEMDIQSKMYSPAPSALTEMSPRAYSSHFEDCN
                     SFNTAQSSPQCFSRFKEYYNGDTLSSYDYPLFPNYMANTQSSKAKARSQSAPKQRPPE
                     IYEKQMSGRRRSSMEAPRNNGVPRAVRMQRSSSQLGSNTAKESQQHHHHQYYPWMAIK
                     LDRSNISLMESECGSTSTVMTNTNYGRHVDVQGNNNMY"
intron               35872. .36296
                     /gene="dl3415w"
                     /number=1
exon                 36297. .36857
                     /gene="dl3415w"
                     /number=2
intron               36858. .36942
                     /gene="dl3415w"
                     /number=2
exon                 36943. .36969
                     /gene="dl3415w"
                     /number=3
exon                 41035. .45039
                     /gene="dl3420w"
                     /number=1
gene                 41035. .46160
                     /gene="dl3420w"
CDS                  join(41035. .45039,45135. .46160)
                     /gene="dl3420w"
                     /note="similarity to centromere protein E (kinesin-related
                     protein CENP-E) - human"
                     /codon_start=1
                     /product="centromere protein homolog"
                     /protein_id="CAB10255.1"
                     /db_xref="GI:2244833"
                     /db_xref="SPTREMBL:O23332"
                     /translation="MDSKVKTMIKLIEADADSFARRADMYFKKRPELMKLVEELYRAY
                     RALAERYDHTTVELRRAHKVMVEAFPNQMSFDMIEDSASSSSEPRTEADTEALQKDGT
                     KSKRSFSQMNKLDGTSDSHEADSEVETLKRTLLELQTEKEALNLQYQLILSKVSRFEK
                     ELNDAQKDVKGFDERACKADIEIKILKESLAKLEVERDTGLLQYSQAIERIADLEASI
                     SHGQEYAKGLTNRVSEAEREAMSLKKELSRLQSEKEAGLLRYNKSLELISSLEKTIRD
                     AEESVRVFRDQSEQAETEIKALKQELLKLNEVNEDLNVRYQQCLETISKLEREVSHAQ
                     DNAKRLSSEVLAGAAKIKTVEEQCALLESFNQTMKVEAENLAHKMSAKDQELSQKQNE
                     IEKLQAVMQEEQLRFSELGASLRNLESLHSQSQEEQKVLTSELHSRIQMLRELEMRNS
                     KLEGDISSKEENRNLSEINDTSISLEIQKNEISCLKKMKEKLEEEVAKQMNQSSALQV
                     EIHCVKGNIDSMNRRYQKLIDQVSLTGFDPESLSYSVKKLQDENSKLVELCTNQRDEN
                     NAVTGKLCEMDSILKRNADLEKLLLESNTKLDGSREKAKDLIERCESLRGEKSELAAE
                     RANLVSQLQIMTANMQTLLEKNSVLEKSLSCANIELESLRDKSKCFDDFFQFLKNDKS
                     ELMKERESLVSQLCKVEEKLGVLEKKYTELEVRYTDLQRDNKLKSHQVEELQVSLAAE
                     KQESANYKRSTESRLADLQKNVSFLREECRSRKREYEDELDRVVNKQVEIFILQKLIE
                     DLEQKNFSLLIECQKHVEASEFSEKLIAELESENLEQQMEAEIFLDEIDSLRGAIYQV
                     IKALQVEADCKTEQKITKDQISVSRALGEIDSLKGSLSSAEYEMHRLVVENSVLLSLL
                     GQFQSDGLVLESEKNILEKDLKTKIHQCGMLEKDKQDLQEANRLLKSKLIKREQQEQK
                     LRAELKFENLKFESLHDSYMVLQQDYSYTLNDNKTLLLKFSEFKDGMHVVEEENDAIL
                     QEAVALSNTCVVYRSFGSEMAEEVEDFVETVSSLREISTGLKRKVETLEKKLEGKEKE
                     SQGLNKMLENLQEGLEEDNFLTGLLEHQVSNVDEILEHREMEILEAEHMLKATNNENE
                     ELHKEVEELRKDYEDSRRMRANLEWQISELSDVAGRQEEEIRKLNALNENLESEVQFL

  Query Match             4.1%;  Score 44.4;  DB 8;  Length 202860;
  Best Local Similarity   49.6%;  Pred. No. 3.7;
  Matches  114;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy     303 gtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaa 362
           || | |   | | |  | | |     ||  | |  |||  |   |    | ||| |   |
Db  136643 GTCTTTGTTAGTTCAAAAAAAAAAGTTTTTGTAATAATAACTCTTTTGAAGTTTGTACGA 136702

Qy     363 gcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatatttt 422
             | ||| |  ||  ||| |||||||| | || |  | | | ||  |  ||  ||| ||  |
Db  136703 ACGTAAGAAAAAAAAAAAAGGAAAACTAGAATTAGTATGACCCAATAACTGAAGCTAGGT 136762

Qy     423 taaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcc 482
           | |   |  | ||  | |||||  ||||  ||||  || ||   | | |   |||   |
Db  136763 TTACCAGGCCGTGTGATAAGAATATAAAGCCCAAATTAAAAATGGGTCTAACCTAAAGGC 136822

Qy     483 tgatgtagctacattaaattctttattcaatcaaataaagaatcagtctt 532
             |   |   | | ||  |||| |   ||  ||  |   | || | | ||
```

```
Db 136823 CCAATAACTGAAACTAGGTTCTCTCCGCATCCATTTTCTGCATTACTATT 136872


RESULT 15
BC016000/c
LOCUS       BC016000               1182 bp    mRNA    linear   PRI 24-OCT-2001
DEFINITION  Homo sapiens, clone IMAGE:4333588, mRNA.
ACCESSION   BC016000
VERSION     BC016000.1  GI:16359070
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1182)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 38 Row: d Column: 13
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .1182
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4333588"
                     /tissue_type="Skin, melanotic melanoma, high MDR."
                     /clone_lib="NIH_MGC_62"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
BASE COUNT      357 a    196 c    195 g    434 t
ORIGIN


  Query Match             4.1%;  Score 44;  DB 9;  Length 1182;
  Best Local Similarity   52.7%;  Pred. No. 7.2;
  Matches  118;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy      18 aacttaacaaaaactcaccattaaactacaaattacaattattatgaaaaatctttttttt 77
           | ||   ||| ||| |    ||  | |||| ||   |||   ||||     | |||||||
Db     289 ATCTATCCAAGAACACGGGGTTTTATTACATATGGAAATCTGTATGGGTTTTTTTTTTTT 230

Qy      78 atcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccgattccaacgggaa 137
            | || | |||  |||| |   || |   || ||  ||  | |   ||| ||    || |
Db     229 TTTAAAGGTGGGTTTTGGGGATGTTTACCCTATTGTTTTTTATTTGGATGCCCTAAGGCA 170

Qy     138 tcaggaaatcaacggaaaggaaaaacta-agtgtaaatgattctaagctgaaagatttcg 196
           |    |   ||  |  ||  |   ||| | |||||||| ||||||  | |||  | |
Db     169 TATAAAGCGCATAGTGAATTACCTACTGGAATGTAAATGGTTCTAAAATTAAAACTGTTA 110

Qy     197 gaaagactgtaccggtagggatagacgaagaaaacggaatgata 240
            ||       |  |  | |||  | | ||||| |  || || |
Db     109 AAATAGAAAGAGGGAAAAGGAAGGCCCAAGAACTCTTAAAGAGA 66
```

Search completed: June 27, 2002, 20:18:48
Job time: 9856 sec



GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 20:21:50 ; Search time 362.26 Seconds
(without alignments)
5118.610 Million cell updates/sec

Title: US-09-727-769A-7
Perfect score: 1080
Sequence: 1 agttaaaataaccaaccaac.....atcagcaggtgctgttatat 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1080	100.0	1080	22	AAF90281	Nucleotide sequenc
	2	555	51.4	555	22	AAF90280	Nucleotide sequenc
	3	527.4	48.8	1080	21	AAZ49495	Chryseobacterium g
	4	344.2	31.9	555	21	AAZ49494	Chryseobacterium g
	5	45.4	4.2	465	21	AAC94109	Cat flea head and
c	6	45.4	4.2	7319	24	ABL34045	Human immune syste
c	7	42.4	3.9	17131	24	ABL33053	Human immune syste
	8	41.8	3.9	1857	21	AAC65441	Staphylococcus war
c	9	41.6	3.9	5987	24	ABL33563	Human immune syste
c	10	41.4	3.8	5884	24	ABL34165	Human immune syste
	11	40.8	3.8	6741	21	AAA10595	Gene encoding a su
c	12	39.6	3.7	19659	24	ABL32767	Human immune syste
	13	39.4	3.6	19070	22	AAK86208	Human immune/haema
c	14	39	3.6	6292	22	AAS46735	Tumour suppressor
	15	38.8	3.6	1565	23	ABL07153	Drosophila melanog
	16	38.8	3.6	2700	23	ABL11892	Drosophila melanog
	17	38.8	3.6	3682	23	ABL07152	Drosophila melanog
c	18	38.8	3.6	5287	24	ABL33493	Human immune syste
c	19	38.6	3.6	9602	22	AAF97874	Human neuroblastom
c	20	38.4	3.6	5121	20	AAX20270	Borrelia burgdorfe
c	21	38.4	3.6	8605	24	ABL32535	Human immune syste
c	22	38.2	3.5	2697	23	ABL13982	Drosophila melanog
c	23	38.2	3.5	2854	23	ABL13954	Drosophila melanog
c	24	38.2	3.5	8103	24	ABL33661	Human immune syste
c	25	38.2	3.5	17293	24	AAS61057	Human gene regulat
	26	37.8	3.5	4590	22	AAH24065	Yeast AOD9604-asso
c	27	37.8	3.5	8513	22	AAS45354	Chemically pretrea
c	28	37.8	3.5	8513	22	AAS46367	Tumour suppressor
	29	37.6	3.5	6071	24	ABL32325	Human immune syste
	30	37.6	3.5	6071	24	AAS61076	Human gene regulat
c	31	37.6	3.5	8093	22	AAS46435	Tumour suppressor
c	32	37.6	3.5	16688	24	ABL32320	Human immune syste
c	33	37.4	3.5	5183	22	AAS45506	Chemically pretrea
c	34	37.4	3.5	5183	24	ABL33349	Human immune syste
c	35	37.4	3.5	8700	24	ABL33002	Human immune syste
c	36	37.4	3.5	13814	24	ABL33193	Human immune syste
	37	37.4	3.5	16818	23	ABL18690	Drosophila melanog
c	38	37.2	3.4	5997	24	ABL33625	Human immune syste
c	39	37.2	3.4	6699	24	AAS63320	Chemically pretrea
c	40	37.2	3.4	40862	24	ABL34073	Human immune syste
c	41	37	3.4	6012	22	AAS46724	Tumour suppressor
c	42	37	3.4	12405	22	AAS45330	Chemically pretrea
c	43	37	3.4	12405	24	AAS61143	Human gene regulat
c	44	37	3.4	15865	22	AAK90825	Human digestive sy
c	45	36.8	3.4	7498	24	ABL32257	Human immune syste

ALIGNMENTS

RESULT 1
AAF90281
ID AAF90281 standard; DNA; 1080 BP.
XX
AC AAF90281;
XX
DT 22-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a protein-deamidating enzyme.
XX
KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW bakery; confectionery; ss.
XX
OS Cryseobacterium sp.
XX
FH Key Location/Qualifiers
FT CDS 61..1023
FT /*tag= a
FT /product= "protein-deamidating enzyme"
XX
PN EP1106696-A1.
XX
PD 13-JUN-2001.
XX
PF 04-DEC-2000; 2000EP-0310768.
XX
PR 03-DEC-1999; 99JP-0345044.
XX
PA (AMAN-) AMANO ENZYME INC.
XX
PI Yamaguchi S;
XX

DR WPI; 2001-376907/40.
DR P-PSDB; AAB84387.
XX
PT New enzyme for use in e.g. bakery has an ability to deamidate amido
PT groups in a protein -
XX
PS Example 11; Page 23; 43pp; English.
XX
CC The present sequence encodes a protein-deamidating enzyme from
CC Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
CC groups in a protein by directly acting upon the amido groups without
CC cutting peptide bonds and without cross-linking the protein. The enzyme
CC thus reduces the mineral sensitivity of the protein and increases the
CC soluble mineral content in the protein-mineral solution, improving the
CC absorption of minerals in the human body. The enzyme reduces or removes
CC the toxicity and allergenic property of the protein in a food
CC e.g. allergy. The enzyme is useful for the improvement of dough in the
CC field of bakery and confectionery e.g. for the production of crackers,
CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC protein products, in various food articles e.g. meat or fish products
CC and noodles; and for improving functionality of plant or animal protein.
XX
SQ Sequence 1080 BP; 371 A; 204 C; 203 G; 302 T; 0 other;

Query Match 100.0%; Score 1080; DB 22; Length 1080;
Best Local Similarity 100.0%; Pred. No. 3.9e-281;
Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 agttaaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattatt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 agttaaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattatt 60

Qy      61 atgaaaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 atgaaaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgt 120

Qy     121 gccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattct 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattct 180

Qy     181 aagctgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgata 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aagctgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgata 240

Qy     241 aaggtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgag 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 aaggtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgag 300

Qy     301 cagtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttctta 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 cagtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttctta 360

Qy     361 aagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 aagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatat 420

Qy     421 tttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaatt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaatt 480

Qy     481 cctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttgcggtacc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 cctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttgcggtacc 540

Qy     541 tctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgttatgcaaga 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 tctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgttatgcaaga 600

Qy     601 gcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttgta 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 gcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatttgta 660

Qy     661 tacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctaccacgttgca 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 tacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctaccacgttgca 720

Qy     721 atattggtaagctataaaaatgcttccggagtaacggaaaaaagaattattgatccttca 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 atattggtaagctataaaaatgcttccggagtaacggaaaaaagaattattgatccttca 780

Qy     781 ctattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgttaacacctct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ctattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgttaacacctct 840

Qy     841 tgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttattacagaagt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 tgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttattacagaagt 900

Qy     901 cctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgtactgactaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 cctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgtactgactaaa 960

Qy     961 ttttcactgctttccggatgttctccttcacctgcaccggatgtatccagctgtggattt 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ttttcactgctttccggatgttctccttcacctgcaccggatgtatccagctgtggattt 1020

Qy    1021 taattaattgataattttacagcacctgctcatttacagaatcagcaggtgctgttatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 taattaattgataattttacagcacctgctcatttacagaatcagcaggtgctgttatat 1080
```

RESULT 2
AAF90280
ID AAF90280 standard; DNA; 555 BP.
XX
AC AAF90280;
XX
DT 22-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a protein-deamidating enzyme.
XX
KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW bakery; confectionery; ss.
XX
OS Cryseobacterium sp.
XX
PN EP1106696-A1.
XX
PD 13-JUN-2001.
XX
PF 04-DEC-2000; 2000EP-0310768.
XX
PR 03-DEC-1999; 99JP-0345044.
XX
PA (AMAN-) AMANO ENZYME INC.
XX
PI Yamaguchi S;
XX
DR WPI; 2001-376907/40.
DR P-PSDB; AAB84386.
XX
PT New enzyme for use in e.g. bakery has an ability to deamidate amido
PT groups in a protein -
XX
PS Claim 8; Page 22; 43pp; English.
XX
CC The present sequence encodes a protein-deamidating enzyme from
CC Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
CC groups in a protein by directly acting upon the amido groups without
CC cutting peptide bonds and without cross-linking the protein. The enzyme
CC thus reduces the mineral sensitivity of the protein and increases the
CC soluble mineral content in the protein-mineral solution, improving the
CC absorption of minerals in the human body. The enzyme reduces or removes
CC the toxicity and allergenic property of the protein in a food
CC e.g. allergy. The enzyme is useful for the improvement of dough in the

```
CC   field of bakery and confectionery e.g. for the production of crackers,
CC   biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC   protein products, in various food articles e.g. meat or fish products
CC   and noodles; and for improving functionality of plant or animal protein.
XX
SQ   Sequence 555 BP; 166 A; 119 C; 107 G; 163 T; 0 other;


  Query Match             51.4%;  Score 555;  DB 22;  Length 555;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-140;
  Matches  555;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    466 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ttggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaat 60

Qy    526 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 cagtcttgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagac 120

Qy    586 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ggatgttatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgt 180

Qy    646 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 705
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gaaaaacaatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtgg 240

Qy    706 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 765
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 agctaccacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaaga 300

Qy    766 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 attattgatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgct 360

Qy    826 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tgcgttaacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaat 420

Qy    886 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gtttattacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaac 480

Qy    946 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 1005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 tgtgtactgactaaattttcactgctttccggatgttctccttcacctgcaccggatgta 540

Qy   1006 tccagctgtggattt 1020
          |||||||||||||||
Db    541 tccagctgtggattt 555


RESULT   3
AAZ49495
ID   AAZ49495 standard; DNA; 1080 BP.
XX
AC   AAZ49495;
XX
DT   04-APR-2000  (first entry)
XX
DE   Chryseobacterium gleum protein-deamidating prepro-enzyme encoding gene.
XX
KW   Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;
KW   amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW   protein engineering; surface hydrophobicity; toxicity; allergic;
KW   mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.
XX
OS   Chryseobacterium gleum 'JCM 2410'.
XX
FH   Key             Location/Qualifiers
FT   CDS             61..1020
FT                   /*tag=  a
FT                   /product= "C. gleum protein deamidating prepro-enzyme"
FT                   /function= "Deamidate amido groups in a protein"
FT   sig_peptide     61..123
FT                   /*tag=  b
FT                   /note= "Corresponds to pre region of the protein"
FT   mat_peptide     463..1017
FT                   /*tag=  c
FT                   /label= Mature_protein_deamidating_enzyme
FT                   /note= "Improves protein function"
XX
PN   EP976829-A2.
XX
PD   02-FEB-2000.
XX
PF   04-JUN-1999;   99EP-0304367.
XX
PR   04-JUN-1998;   98JP-0173940.
XX
PA   (AMAN ) AMANO PHARM KK.
XX
PI   Yamaguchi S,  Matsuura A;
XX
DR   WPI; 2000-118552/11.
DR   P-PSDB; AAY44583.
XX
PT   New enzyme for modifying and improving the function of proteins and/or
PT   peptides has deamidating activity without causing cross linking -
XX
PS   Example 26; Page 25; 57pp; English.
XX
CC   The present sequence is the gene encoding the protein-deamidating
CC   prepro-enzyme, isolated from a new strain of soil bacterium,
CC   Chryseobacterium gleum JCM 2410. The enzyme exerts the deamidation
CC   activity by directly acting upon side chain amido groups in the protein
CC   in bonded state and releasing side chain carboxyl groups and ammonia. It
CC   can deaminate high molecular weight proteins, without cross linking and
CC   cleavage of peptide bonds, to improve protein function. This sequence is
CC   used for protein engineering, to cause an increase in surface
CC   hydrophobicity and improve the function of a plant or animal protein. It
CC   can also be used to remove or reduce toxicity or allergic property of
CC   proteins in food, decrease mineral sensitivity of protein, to allow
CC   greater absorption into the body and to solubilise calcium for use in
CC   drinks and mineral enhancing agents.
XX
SQ   Sequence 1080 BP; 354 A; 210 C; 205 G; 311 T; 0 other;


  Query Match             48.8%;  Score 527.4;  DB 21;  Length 1080;
  Best Local Similarity   70.9%;  Pred. No. 3.4e-132;
  Matches  715;  Conservative    0;  Mismatches  291;  Indels    3;  Gaps    1;

Qy     17 caacttaacaaaaactcaccattaaactacaaattacaattattatgaaaaatctttttt 76
          ||| | || |       |  | |||||  |||| || ||  | |||||||||  || | |
Db     17 caattaaaaagttcactaaaactaaacaccaaaatataaaaactatgaaaaaatttctgt 76

Qy     77 tatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccgattccaacggga 136
          |||| |||||||| || |||||  |  |  | || ||| ||||  | |||||  |  |  |
Db     77 tatccatgatggcattcgtgacgattctgtcattcaatgcctgctcagattcaagtgcca 136

Qy    137 atcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagatttcg 196
          | |||||    ||       |  |||  |  |    | |   |||   ||||||||||||
Db    137 accaggacccgaatcttgtcgctaaagaatctaacgaagtcgcta---tgaaagatttcg 193

Qy    197 gaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatttatgt 256
          | |||||||| ||||||||||| ||  ||||| | ||||   | ||  | |||||||||
Db    194 gtaagactgttccggtagggattgaaaaagaagatggaaaatttaaaatctcatttatgg 253

Qy    257 taactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaatgc 316
          | ||||| ||    |||||||||  |   |  ||||||||||     |||||    |||
Db    254 ttactgcccagccgtatgaaattgcggacagtaaagaaaatgcaggttatatttccatga 313

Qy    317 ttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaatg 376
          | ||||||||||||| |||||||| |||| || ||  ||||| | ||    || | ||||
```

```
Db    314 tcagacaggctgttgagaatgaaactcccgttcatgttttccttaaagtcaacaccaata 373

Qy    377 aaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcctga 436
          |||| | |||||| ||    ||||     || ||| |  | |||||||||| |  |  | |
Db    374 aaattgcaaaagtagaaaaagcaacagatgatgacatccgttatttttaaatctgtattca 433

Qy    437 caaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctgatgtagctacat 496
            ||  |||  | |||  ||| ||| |||   |  ||||| |||||||||| | || ||
Db    434 acaagcaagagagaggtgaaagcaacaaagcagtcagtgttattcctgatctggcaacgc 493

Qy    497 taaattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctcac 556
          | ||   |||||| |  || || || || ||| ||||||| || ||||| || || || |
Db    494 tgaacagtttatttacccagatcaaaaaccaggcttgcggaacttctacagcatcttctc 553

Qy    557 catgcatcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagac 616
          | || ||||| |||||||||| || ||||||||||||||||| || || || |||||||
Db    554 cttgtatcaccttcagatatccggttgacggatgttatgcaagggctcacaaaatgagac 613

Qy    617 aaatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaagg 676
          ||||| || |||||  |||||||||||||||||| || || ||||| || || || | |
Db    614 aaatcctattgaacgccggctatgactgtgaaaagcagttcgtatatggtaatctgagag 673

Qy    677 catcaacaggaacttgctgtgtggcgtggagctaccacgttgcaatattggtaagctata 736
          | || |||||||| ||||||||  | |||   || ||||| ||||| ||||||||||  |
Db    674 cttctacaggaacatgctgtgtatcatgggtatatcacgtagcaattttggtaagcttca 733

Qy    737 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 796
          |||||||||| ||| |    |||||||||| || |||||||| |||| || |||||||
Db    734 aaaatgcttcaggaattgttgaaaaaagaatcatagatccttcattattctccagcggtc 793

Qy    797 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 856
          |||||||||| | |||||||||   || ||    |||||    |||||||||| || |
Db    794 ctgtaacagattctgcatggagagctgcatgtaccaacacaagctgcggatctgcgtctg 853

Qy    857 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 916
          | || || || || ||||| |||||||||||||| ||||||||||      ||| |  |
Db    854 tatcttcctacgccaatacagcaggaaatgtttactacagaagtccgtcaggttcattac 913

Qy    917 tgtatgacaacaatctgatcaataccaactgtgtactgactaaattttcactgctttccg 976
          ||||||| |||||     | |||||||| |||||| | |  | ||| |||   |||||
Db    914 tgtatgataacaactatgtgaataccaattgtgtattaaacatattctcatccctttcag 973

Qy    977 gatgttctccttcacctgcaccggatgtatccagctgtggattttaatt 1025
          ||||||||||||| || |||||   |||| | ||||||||||||||||
Db    974 gatgttctccttccccagcaccaagtgtagcaagctgtggattttaatt 1022
```

RESULT 4
AAZ49494
ID AAZ49494 standard; DNA; 555 BP.
XX
AC AAZ49494;
XX
DT 04-APR-2000 (first entry)
XX
DE Chryseobacterium gleum protein-deamidating enzyme encoding DNA.
XX
KW Protein deamidating enzyme; soil bacterium; deamidation activity;
KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW protein engineering; surface hydrophobicity; toxicity; allergic;
KW mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.
XX
OS Chryseobacterium gleum 'JCM 2410'.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..555
FT /*tag= a
FT /product= "Protein deamidating enzyme"
FT /function= "Deamidate amido groups in a protein"
FT /note= "Improves protein function"
XX
PN EP976829-A2.
XX
PD 02-FEB-2000.
XX
PF 04-JUN-1999; 99EP-0304367.
XX
PR 04-JUN-1998; 98JP-0173940.
XX
PA (AMAN) AMANO PHARM KK.
XX
PI Yamaguchi S, Matsuura A;
XX
DR WPI; 2000-118552/11.
DR P-PSDB; AAY44582.
XX
PT New enzyme for modifying and improving the function of proteins and/or
PT peptides has deamidating activity without causing cross linking -
XX
PS Claim 8; Page 23; 57pp; English.
XX
CC The present sequence is the DNA encoding the protein-deamidating enzyme,
CC isolated from a new strain of soil bacterium, Chryseobacterium gleum
CC JCM 2410. The enzyme exerts the deamidation activity by directly acting
CC upon side chain amido groups in the protein in bonded state and
CC releasing side chain carboxyl groups and ammonia. It can deaminate high
CC molecular weight proteins, without cross linking and cleavage of peptide
CC bonds, to improve protein function. This sequence is used for protein
CC engineering, to cause an increase in surface hydrophobicity and improve
CC the function of a plant or animal protein. It can also be used to remove
CC or reduce toxicity or allergic property of proteins in food, decrease
CC mineral sensitivity of protein, to allow greater absorption into the body
CC and to solubilise calcium for use in drinks and mineral enhancing agents.
XX
SQ Sequence 555 BP; 159 A; 122 C; 112 G; 162 T; 0 other;

Query Match 31.9%; Score 344.2; DB 21; Length 555;
Best Local Similarity 76.7%; Pred. No. 6.5e-83;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy    472 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 531
          ||||| |||||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db      7 agtgttattcctgatctggcaacgctgaacagtttatttacccagatcaaaaaccaggct 66

Qy    532 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 591
          ||||| || ||||| || || || || || ||||| |||||||||| || |||||||||
Db     67 tgcggaacttctacagcatcttctccttgtatcaccttcagatatccggttgacggatgt 126

Qy    592 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 651
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    127 tatgcaagggctcacaaaatgagacaaatcctattgaacgccggctatgactgtgaaaag 186

Qy    652 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 711
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    187 cagttcgtatatggtaatctgagagcttctacaggaacatgctgtgtatcatgggtatat 246

Qy    712 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 771
          ||||| ||||| ||||||||||  |||||||||||| ||| |    |||||||||| ||
Db    247 cacgtagcaattttggtaagcttcaaaaatgcttcaggaattgttgaaaaaagaatcata 306

Qy    772 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 831
          ||||||||| |||| || |||||||||||||||||| | |||||||||   || ||
Db    307 gatccttcattattctccagcggtcctgtaacagattctgcatggagagctgcatgtacc 366

Qy    832 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 891
          |||||    |||||||||||| || || || || || || ||||| ||||||||||||
Db    367 aacacaagctgcggatctgcgtctgtatcttcctacgccaatacagcaggaaatgtttac 426

Qy    892 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 951
          ||||||||||      ||| |  |||||||| |||||     | |||||||| ||||||
Db    427 tacagaagtccgtcaggttcattactgtatgataacaactatgtgaataccaattgtgta 486

Qy    952 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 1011
           | |  | ||| |||   ||||| |||||||||||||| || |||||   |||| | |||
Db    487 ttaaacatattctcatccctttcaggatgttctccttccccagcaccaagtgtagcaagc 546
```

```
Qy   1012 tgtggattt 1020
          |||||||||
Db    547 tgtggattt 555
```

RESULT 5
AAC94109
ID AAC94109 standard; cDNA; 465 BP.
XX
AC AAC94109;
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:604.
XX
KW Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
KW vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
XX
OS Ctenocephalides felis.
XX
PN WO200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09437.
XX
PR 09-APR-1999; 99US-0128704.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisnewski N;
XX
DR WPI; 2000-656323/63.
XX
PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations -
XX
PS Claim 26; Page 444; 964pp; English.
XX
CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
CC anti-HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HNC cDNA of the invention.
XX
SQ Sequence 465 BP; 166 A; 69 C; 63 G; 167 T; 0 other;

Query Match 4.2%; Score 45.4; DB 21; Length 465;
Best Local Similarity 51.2%; Pred. No. 0.021;

Matches 106; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```
Qy    210 ggtagggatagacgaagaaaacggaatgataaaggtgtcatttatgttaactgcgcaatt 269
          | || | ||  |    ||| |   ||   |  ||||||| |||||||||  | |  | ||
Db     94 gctacgtatctaaattgaataaaaaagcgtgtaggtgtcttttatgttacatacttattt 153

Qy    270 ctatgaaattaagccgaccaaagaaaatgagcagtatatcggaatgcttagacaggctgt 329
             | || ||    | |  | |||| || |  | ||  |  | ||  ||||  |   | |
Db    154 gattaaatttcctgcaaatatagaatattaataataatttcggatcattagtaattttat 213

Qy    330 taagaatgaatctcctgtacacattttcttaaagcctaatagcaatgaaataggaaaagt 389
          |||  || |   |     | | |     || |   | | |    | | ||  ||||| |
Db    214 taaatattatcatttacaatatacaaaatttacatattacatagttaacattagaaaact 273

Qy    390 ggagtctgcaagtccggaagacgtaag 416
            | | ||  |||||  || | || ||
Db    274 aaaatttgatagtccataaaatgttag 300
```

RESULT 6
ABL34045/c
ID ABL34045 standard; DNA; 7319 BP.
XX
AC ABL34045;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2018.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 2018; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 7319 BP; 1946 A; 263 C; 1552 G; 3558 T; 0 other;

```
  Query Match             4.2%;  Score 45.4;  DB 24;  Length 7319;
  Best Local Similarity   46.7%;  Pred. No. 0.058;
  Matches  179;  Conservative    0;  Mismatches  201;  Indels    3;  Gaps    1;

Qy       5 aaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatga 64
           ||| |||  ||   | || | | |||  ||    || |  |  |||||||| | ||| |
Db    6212 AAACTAATAAAATTAATTCAAACAAATCCAAATATATATCAATAATTACAATAAATATAA 6153

Qy      65 aaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccg 124
           | ||  |    ||| |||| |      | |   ||| | |||| |   ||  |     ||
Db    6152 ATAAAATAAACTTACCAATTAAAAATTATACTACCATATTAAATAAAAAAACAAAATCCA 6093

Qy     125 attccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagc 184
           |   | |||  ||      ||   ||  ||  |||||  ||    |||| |    ||
Db    6092 A---CTACGTAAAATTTATAAAACACAAAATAAAAAAAACAAAAAAAAATTAAAAAAAAA 6036

Qy     185 tgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaagg 244
             ||| |      ||| |    |     |     | |  || |||||  ||  || |
Db    6035 ATAAAAAACGAAAAAATATATAAAATAAACACCAATAAAAAAAAAACTAAAATATTACCT 5976

Qy     245 tgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagt 304
           |   |||||    || |    ||    |  | | ||     |   || ||||  |  | |
Db    5975 TAATATTTAACAAAAATTTTAAAACAAAAAACAATATTAATAATCAAAAAAACAACTACT 5916

Qy     305 atatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagc 364
           | ||   ||   || |      || ||  ||| |       | ||||      |
Db    5915 AAATAATAAAAAATAAATTCATCAAAAAAAAAAAATATAACAATTTAAATTTCCATTTTC 5856

Qy     365 ctaatagcaatgaaataggaaaa 387
           |||||| |    ||||| ||||
Db    5855 CTAATAACCTCAAAATATAAAAA 5833


RESULT   7
ABL33053/c
ID   ABL33053 standard; DNA; 17131 BP.
XX
AC   ABL33053;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1026.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 1026; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 17131 BP; 5134 A; 253 C; 3328 G; 8413 T; 3 other;


  Query Match             3.9%;  Score 42.4;  DB 24;  Length 17131;
  Best Local Similarity   52.9%;  Pred. No. 0.51;
  Matches   91;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy      23 aacaaaaactcaccattaaactacaaattacaattattatgaaaaatctttttttatcaa 82
           ||| | ||| ||  |   || | ||| ||||||  ||| | | ||||    ||  || ||
Db    1443 AACTATAACCCATAAACCAAATCCAACTTACAACCATTTTTATAAATAAAATTACATAAA 1384

Qy      83 tgatggcctttgtgaccgtcttaacttttaattcctgtgccgattccaacgggaatcagg 142
              ||| |     | | || || | | ||   |       ||   ||   ||| | |
Db    1383 AATATACCTATACCCACTTATTTACATATTATATATAACTACTTTTACACTTTAATAACG 1324

Qy     143 aaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagattt 194
           |||| |||  |   |||||  || |  |||       || || ||| ||||
Db    1323 AAATTAACTAATTACAAAAAAAAACTTAAAACCCACGAAATCTAAAATATTT 1272


RESULT   8
AAC65441
ID   AAC65441 standard; DNA; 1857 BP.
XX
AC   AAC65441;
XX
DT   14-FEB-2001  (first entry)
XX
DE   Staphylococcus warneri SwaI methylase gene.
XX
KW   Staphylococcus warneri; SwaI restriction-modification gene;
KW   restriction endonuclease; SwaI methylase; ds.
XX
OS   Staphylococcus warneri.
XX
PN   EP1048731-A2.
XX
PD   02-NOV-2000.
XX
PF   26-APR-2000; 2000EP-0303507.
XX
PR   27-APR-1999;   99US-0299378.
XX
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI   Kong H,  Higgins LS,  Dalton MA;
XX
DR   WPI; 2000-681207/67.
DR   P-PSDB; AAB28685.
XX
PT   New method for cloning and producing the SwaI restriction endonuclease
PT   from Staphylococcus warneri which can be produced in abundance from
PT   Escherichia coli  -
XX
PS   Disclosure; Fig 3; 18pp; English.
XX
CC   The present sequence encodes the Staphylococcus warneri SwaI
CC   methylase of the SwaI restriction-modification system. SwaI restriction
CC   endonucleases are used for characterising genes in the laboratory. The
CC   bacterial strains used permit better and simplified purification methods.
XX
SQ   Sequence 1857 BP; 793 A; 203 C; 220 G; 641 T; 0 other;
```

```
  Query Match             3.9%;  Score 41.8;  DB 21;  Length 1857;
  Best Local Similarity   48.9%;  Pred. No. 0.33;
  Matches  112;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy     302 agtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaa 361
           || |  | |  ||   |  | | ||| |||   ||| |  |  || ||| ||  || |
Db     905 agcaacttgatatcaatttaaaagcttttagtgatgtagttagtggacaaataatcattt 964

Qy     362 agcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatatt 421
             | |   |  |||| |     ||||  |    || ||||    ||||       | ||
Db     965 ctcttttaaaaaatgcatccaacaaagaagctattgtaagtatcaaagattggcaaaata 1024

Qy     422 ttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattc 481
            ||| || || |  |  ||  |||  | | ||       |||||||    || | | |
Db    1025 ataacacaattcaaattaatcaagatatatggttgcatgataaatttttatcgttttaata 1084

Qy     482 ctgatgtagctacattaaattctttattcaatcaaataaagaatcagtc 530
           |   |     | | ||||||  | ||  |  || |  | ||| | ||
Db    1085 ttagcgataagaaaataaattcaattttagagaaagtttacaataaatc 1133


RESULT   9
ABL33563/c
ID   ABL33563 standard; DNA; 5987 BP.
XX
AC   ABL33563;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1536.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 1536; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 5987 BP; 1928 A; 96 C; 1209 G; 2754 T; 0 other;


  Query Match             3.9%;  Score 41.6;  DB 24;  Length 5987;
  Best Local Similarity   46.5%;  Pred. No. 0.57;
  Matches  134;  Conservative    0;  Mismatches  154;  Indels    0;  Gaps    0;

Qy       6 aaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatgaa 65
           |||| | ||    | | ||  ||||   |      |  || ||||||    ||||   ||
Db     534 AAATTATCATTATAATAAAACAAAATCTAATCAATATATATAAATTATTTATATTTAAAA 475

Qy      66 aaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccga 125
             | || ||  || ||  |     |||   |   ||| | |  ||  |||||   || |
Db     474 CCTTTTTATTAAATAAAAAAAATTATTTTAAATATTCTAATCAATTCTTTCCTACTCCCA 415

Qy     126 ttccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagct 185
             |||  |||| |     | ||  ||  || | || |     ||    | | |  |
Db     414 AAAAAACCCGAATTATATTTTAAAATAAACAAAATATCTTAAAAATAAACTATTTTACTT 355

Qy     186 gaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggt 245
            |||  ||||   ||   |  |   ||     |  | |    ||   ||  |  ||  |
Db     354 TAAATTTTTCCCTAAATATTAATTAATATACTAAATCCATTTTAATTTAAAAAATAAAAT 295

Qy     246 gtcatttatgttaactgcgcaattctatgaaattaagccgaccaaaga 293
           | |||| | |||| |  ||   ||  || || | |     ||| |
Db     294 ACTACTTATTTAAACTTCTTAAAAATACTAATTTTAACATCTAAAAAA 247


RESULT  10
ABL34165/c
ID   ABL34165 standard; DNA; 5884 BP.
XX
AC   ABL34165;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2138.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2138; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
```

CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5884 BP; 1655 A; 98 C; 1269 G; 2862 T; 0 other;

Query Match 3.8%; Score 41.4; DB 24; Length 5884;
Best Local Similarity 46.0%; Pred. No. 0.64;
Matches 177; Conservative 0; Mismatches 206; Indels 2; Gaps 1;

```
Qy       5 aaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatga 64
Db    3367 AAAATAATCAAATATTTAATTAAAACAAATAAAAAATAATATACATCATATAAAATATAA 3308

Qy      65 aaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccg 124
Db    3307 ATTTATATATAAAATCTTTAATCCAAATTATTACCCTCCTACCCTTCTAAATAACCGTCA 3248

Qy     125 attccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagc 184
Db    3247 TATCCTAATCTAAAAATCAAATAAACTAAAAATAAAATTCAACCAAACAAAATTCTATAA 3188

Qy     185 tgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaagg 244
Db    3187 TACTTTCATCACACACAAATTTATTACTTTTTACCTACTCTCAAAATAAAATAATAAATC 3128

Qy     245 tgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagt 304
Db    3127 TCTAAAAAAAAATTAATTTACT--TCATAAAAAAAAAAACCAACTAAAAATTATTTCCTAT 3070

Qy     305 atatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagc 364
Db    3069 AACCACTAACTAATACTAAAAAACATTAAAAATATTCCCATTAAAAAAATATCTACAAAT 3010

Qy     365 ctaatagcaatgaaataggaaaagt 389
Db    3009 AAAAAAAAATTAAACAAAAAAAAAT 2985
```

RESULT 11
AAA10595
ID AAA10595 standard; DNA; 6741 BP.
XX
AC AAA10595;
XX
DT 29-JUN-2000 (first entry)
XX
DE Gene encoding a subunit of cellulose synthase.
XX
KW Cellulose synthase; cellulose production; increase yield; ds.
XX
OS Vigna angularis.
XX
PN JP2000060568-A.
XX
PD 29-FEB-2000.
XX
PF 26-AUG-1998; 98JP-0239998.
XX
PR 26-AUG-1998; 98JP-0239998.
XX
PA (MIZU/) MIZUNO K.
PA (OJIP) OJI PAPER CO.
XX
DR WPI; 2000-342371/30.
DR P-PSDB; AAY85180.
XX
PT A gene encoding a cellulose synthetic equipment - for the improvement
PT in the amount of cellulose synthesised in a plant body
XX
PS Claim 2; Page 22-31; 32pp; Japanese.
XX
CC This sequence represents a gene encoding a subunit of the cellulose
CC synthase complex of Vigna angularis. The invention relates to subunits of
CC cellulose synthetic equipment, that can be used to increase the amount of
CC cellulose synthesised by a plant. The proteins and genes encoding them
CC can also be used to improve the properties of the cellulose being
CC produced by a plant.
XX
SQ Sequence 6741 BP; 1712 A; 870 C; 1468 G; 1374 T; 1317 other;

Query Match 3.8%; Score 40.8; DB 21; Length 6741;
Best Local Similarity 24.4%; Pred. No. 0.98;
Matches 92; Conservative 109; Mismatches 175; Indels 1; Gaps 1;

```
Qy      40 aaactacaaattacaattattatgaaaaatctttttttatcaatgatggcctttgtgacc 99
Db    5952 aayshasnasnvahvaysasnaaassrthrsrasasgysthrhgyghgythrthrsraav 6011

Qy     100 gtcttaacttttaattcctgtgccgattccaacgggaatcaggaaatcaacggaaaggaa 159
Db    6012 avamtargasgyasgyyssryscyshgyhvaasnhgasnaaasasaaaaargaavagaaa 6071

Qy     160 aaactaagtgtaaatgattctaagctgaaagatttcggaaagactgtaccggtagggata 219
Db    6072 sngyysyshasasysgtrtyrvagyysaagnysyssrgarggasngysgnarghggnsrm 6131

Qy     220 gacgaagaaaacggaatgataaaggtgtcatttatgttaactgcgcaattctatgaaatt 279
Db    6132 tysgaa-aaasystyrgngyaaasntyrvaysasnasassrsrasasysysghsrrhgyt 6190

Qy     280 aagccgaccaaagaaaatgagcagtatatcggaatgcttagacaggctgttaagaatgaa 339
Db    6191 hrthrsrcysysvamtargasrasngyvasrarggysrgyhvaaahsrthrrggaasrar 6250

Qy     340 tctcctgtacacattttcttaaagcctaatagcaatgaaataggaaaagtggagtctgca 399
Db    6251 gaasrgmtasngyysmtvaaasrysrtyrvathraagnargysgasargargaaarggna 6310

Qy     400 agtccggaagacgtaag 416
Db    6311 agnhaagnmtargrvag 6327
```

RESULT 12
ABL32767/c
ID ABL32767 standard; DNA; 19659 BP.
XX
AC ABL32767;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 740.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX

```
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 740; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 19659 BP; 6577 A; 68 C; 3643 G; 9371 T; 0 other;


  Query Match             3.7%;  Score 39.6;  DB 24;  Length 19659;
  Best Local Similarity   58.5%;  Pred. No. 3.1;
  Matches   69;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy       5 aaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatga 64
           ||||    |  |||   ||  ||  |||   || |  | ||  |||| |  ||| |||
Db   11082 AAAAATTTCTCCCATTCTATAAATTACTATTCACTCTAATAATAATTTCTTTTACTATAC 11023

Qy      65 aaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgc 122
           |||| ||||||   | ||| |   ||  | |  |  | ||||||||||| | ||| | |
Db   11022 AAAAACTTTTTAATTTAATTAAATCCCATTTATCAATTTTAACTTTTATTACCTTTAC 10965


RESULT  13
AAK86208
ID   AAK86208 standard; DNA; 19070 BP.
XX
AC   AAK86208;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41020.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
```

PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 41020; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 19070 BP; 5077 A; 4466 C; 4474 G; 5053 T; 0 other;

Query Match 3.6%; Score 39.4; DB 22; Length 19070;
Best Local Similarity 49.8%; Pred. No. 3.4;
Matches 100; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 325 gctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaatgaaatagga 384
Db 18683 gatgtaataattatatgtaagggcaaaatttattttagatataaagtaaaggaggagggt 18742

Qy 385 aaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcctgacaaaagaa 444
Db 18743 gagatgctttctgcatttcttgatgacagtttgtgttcctacaaaaataaacaaaatgaa 18802

Qy 445 gtaaaagggcaaaccaataaattggcgagtgtaattcctgatgtagctacattaaattct 504
Db 18803 taaaaagggatcaccattcaatttgagtttccaggggaagtgcatgtataatgaaatga 18862

Qy 505 ttattcaatcaaataaagaat 525
Db 18863 taatggacttcaatgaagaat 18883

RESULT 14
AAS46735/c
ID AAS46735 standard; DNA; 6292 BP.
XX
AC AAS46735;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #459.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
PS Claim 1; SEQ ID No 459; 27pp; English.

```
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6292 BP; 1736 A; 99 C; 1304 G; 3153 T; 0 other;


  Query Match             3.6%;  Score 39;  DB 22;  Length 6292;
  Best Local Similarity   42.2%;  Pred. No. 2.9;
  Matches  219;  Conservative    0;  Mismatches  300;  Indels    0;  Gaps    0;

Qy       5 aaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatga 64
           |||||||  ||  | | ||| |||||  |   | ||||| | |||  | ||  |||   |
Db    4352 AAAATAAAAAAAAATCGTAAAAAAAAAACGAAAATAAACAAAAAAAAAAAAAAAATTTACA 4293

Qy      65 aaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccg 124
           ||||| |  |   |  ||  |           |   | | ||     || |  |
Db    4292 AAAATATAATAAAAAAAAAAAAAAAAAAAAAAAAAACTATAAATAAAAAAATTTTAAATAA 4233

Qy     125 attccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagc 184
             ||    ||  |  ||   |    |||  |||||  ||    |||  |    ||
Db    4232 TAAAAAAAAAAAAAAAATAACAAATAAAAAAAAAAAAAAAAAATACAAAAAAAAAAAAAAA 4173

Qy     185 tgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaagg 244
             ||| | |  |||  |   |    |   |    || || |   ||  | |||
Db    4172 AAAAAAAATAACGAACAAAAAAAAAAAAAAAATTTTATAAAAAAACAATAAAAAAAAAAAC 4113

Qy     245 tgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagt 304
           | | |  ||  | ||      ||    |  ||   |     |  ||| ||||  |  |
Db    4112 TTTTAAATAAATAAAAAAAAAAAAAAAAAAATAACCAATAAAAAAAAAATAAAAAAAATAAC 4053

Qy     305 atatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagc 364
             | || ||     | | |   |  ||| ||   ||   |   | | |   | | |||
Db    4052 GAACCGAAAAAAAAAAAAAATTTTACTAAAAAACGATAAACAAAAAAAAAAAATATAAAAAA 3993

Qy     365 ctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatatttta 424
             || |  |     | |  | |  |  |      ||      || |   |  | |  | |
Db    3992 AAAAAACAATAACTAAAAAATATTTAAATAAACGAAAAAAAAAAAAAAAATAACAAATAA 3933

Qy     425 aaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctg 484
           |||  |     | || | ||  ||||     ||  || |||      |    ||    |
Db    3932 AAAAAAAATAAAAAATATAACCAAAAAAAATAAAAAAAAAAAACTAAAAAAAAAACGATA 3873

Qy     485 atgtagctacattaaattctttattcaatcaaataaaga 523
           |   |  ||  ||||| |   ||   ||  ||| ||| |
Db    3872 AAAAAAATAATTTAAAATAAATAAAAAACAAAAAAAAAA 3834


RESULT  15
ABL07153
ID   ABL07153 standard; cDNA; 1565 BP.
XX
AC   ABL07153;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 15941.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB63050.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 15941; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1565 BP; 521 A; 376 C; 361 G; 307 T; 0 other;


  Query Match             3.6%;  Score 38.8;  DB 23;  Length 1565;
  Best Local Similarity   45.0%;  Pred. No. 2;
  Matches  145;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy     136 aatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagatttc 195
           ||   | ||  ||| | |||| | |||  | ||||   |  |  |||    || ||
Db     690 aaagggaaagccaatgaaaagaagaaaaaacgtgttggtactgataaagcaaacgaagct 749

Qy     196 ggaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatttatg 255
             |||     |           |||  || || | |     | ||  |||  |   ||
Db     750 ttgaagcagaaagaaactaatgaagataaaaaagaagaggaagttaatttgtgctccctg 809

Qy     256 ttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaatg 315
             || |||  ||||    | || |   |  |     ||  |  |||| | ||  | ||
Db     810 ggaaatgctaaatttggaggaactggacaaatgccggagtaccagcactttaaggcaaaa 869

Qy     316 cttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaat 375
             ||  | ||   | |||   |   |||    | |      | |    | |  ||||  |
Db     870 gttcaaaagaaggataatgctttgtctagctttcttggaatgtacgggtcagatagtgac 929

Qy     376 gaaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcctg 435
           |||  | |  |||  |||  ||    || | |    | |   |    ||||   ||||
Db     930 gaagaaagtgaagctgagagtgatgatcagaatacagcagactgcagtaaacaaatccca 989

Qy     436 acaaaagaagtaaaagggcaaa 457
```

```
              || ||| ||  |     ||||
Db    990 gcacaagtagagagctcccaaa 1011
```

Search completed: June 27, 2002, 20:22:25
Job time: 6443 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 20:15:14 ; Search time 103.12 Seconds
(without alignments)
2572.577 Million cell updates/sec

Title: US-09-727-769A-7
Perfect score: 1080
Sequence: 1 agttaaaataaccaaccaac.....atcagcaggtgctgttatat 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	527.4	48.8	1080	4	US-09-324-910-11	Sequence 11, Appl
	2	344.2	31.9	555	4	US-09-324-910-5	Sequence 5, Appli
c	3	60	5.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	4	41.8	3.9	1857	4	US-09-299-378-3	Sequence 3, Appli
c	5	35.6	3.3	2489	3	US-09-141-047-7	Sequence 7, Appli
c	6	35.4	3.3	1215	2	US-09-092-770-8	Sequence 8, Appli
c	7	35.4	3.3	1215	4	US-09-222-851-8	Sequence 8, Appli
	8	34.8	3.2	1751	4	US-09-004-393B-3	Sequence 3, Appli
c	9	34.6	3.2	1212	2	US-09-092-770-18	Sequence 18, Appl
c	10	34.6	3.2	1212	4	US-09-222-851-18	Sequence 18, Appl
	11	34.2	3.2	1723	1	US-08-417-492-1	Sequence 1, Appli
	12	34.2	3.2	3666	2	US-08-682-517-13	Sequence 13, Appl
	13	34.2	3.2	3666	2	US-08-682-517-14	Sequence 14, Appl
	14	34.2	3.2	4197	2	US-08-682-517-7	Sequence 7, Appli
	15	34.2	3.2	4197	2	US-08-682-517-8	Sequence 8, Appli
c	16	33.6	3.1	4223	4	US-09-541-782-5	Sequence 5, Appli
	17	33.2	3.1	1198	1	US-08-020-245A-3	Sequence 3, Appli
	18	33.2	3.1	1198	1	US-08-396-957A-3	Sequence 3, Appli
	19	33.2	3.1	2304	1	US-08-020-245A-1	Sequence 1, Appli
	20	33.2	3.1	2304	1	US-08-396-957A-1	Sequence 1, Appli
	21	33	3.1	289	4	US-09-007-005-17	Sequence 17, Appl
	22	33	3.1	289	4	US-09-244-796-17	Sequence 17, Appl
	23	33	3.1	8050	4	US-09-491-362-11	Sequence 11, Appl
	24	32.8	3.0	4002	3	US-09-356-952-9	Sequence 9, Appli
c	25	32.8	3.0	5502	3	US-08-836-134-1	Sequence 1, Appli
	26	32.4	3.0	7286	3	US-09-331-581-3	Sequence 3, Appli
	27	32.4	3.0	7938	3	US-09-331-581-14	Sequence 14, Appl
	28	32	3.0	3095	6	5231168-1	Patent No. 5231168
	29	31.8	2.9	1189	4	US-08-936-165A-195	Sequence 195, App
c	30	31.8	2.9	4471	2	US-08-615-942A-1	Sequence 1, Appli
	31	31.8	2.9	7218	1	US-08-232-463-14	Sequence 14, Appl
	32	31.6	2.9	1810	3	US-08-946-026-19	Sequence 19, Appl
	33	31.6	2.9	1820	4	US-08-845-258-17	Sequence 17, Appl
c	34	31.6	2.9	1820	4	US-08-845-258-37	Sequence 37, Appl
	35	31.6	2.9	1820	4	US-08-990-571-17	Sequence 17, Appl
c	36	31.6	2.9	1820	4	US-08-990-571-37	Sequence 37, Appl
	37	31.6	2.9	1820	4	US-08-723-142A-17	Sequence 17, Appl
c	38	31.6	2.9	1820	4	US-08-723-142A-37	Sequence 37, Appl
c	39	31.4	2.9	2747	2	US-08-874-347-1	Sequence 1, Appli
c	40	31.4	2.9	2747	3	US-09-093-522-1	Sequence 1, Appli
c	41	31.4	2.9	10614	1	US-08-135-511-35	Sequence 35, Appl
c	42	31.4	2.9	10614	1	US-08-187-453-35	Sequence 35, Appl
	43	31.2	2.9	657	3	US-08-721-979A-74	Sequence 74, Appl
	44	31.2	2.9	657	3	US-08-836-501-74	Sequence 74, Appl
	45	31.2	2.9	726	3	US-08-836-501-78	Sequence 78, Appl

ALIGNMENTS

RESULT 1
US-09-324-910-11
; Sequence 11, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Chryseobacterium gleum
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (61)..(462)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (463)..(1017)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (61)..(1017)
US-09-324-910-11

Query Match 48.8%; Score 527.4; DB 4; Length 1080;
Best Local Similarity 70.9%; Pred. No. 7.1e-144;
Matches 715; Conservative 0; Mismatches 291; Indels 3; Gaps 1;

```
Qy      17 caacttaacaaaaactcaccattaaactacaaattacaattattatgaaaaatctttttt 76
           ||| | || |       |  | |||||  |||| || ||  | |||||||||  || | |
Db      17 caattaaaaagttcactaaaactaaacaccaaaatataaaaactatgaaaaaatttctgt 76

Qy      77 tatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccgattccaacggga 136
           |||| |||||||| || |||||  |  |  | || ||| ||||  | ||||| |  |  |
Db      77 tatccatgatggcattcgtgacgattctgtcattcaatgcctgctcagattcaagtgcca 136

Qy     137 atcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagatttcg 196
           | |||||    ||       |  |||  | |    | |   |||   ||||||||||||
Db     137 accaggacccgaatcttgtcgctaaagaatctaacgaagtcgcta---tgaaagatttcg 193

Qy     197 gaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatttatgt 256
           | |||||||| |||||||||| ||  ||||| | ||||   | ||  | |||||||||
```

```
Db    194 gtaagactgttccggtagggattgaaaaagaagatggaaaatttaaaatctcatttatgg 253

Qy    257 taactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaatgc 316
          | ||||| ||    |||||||||  |   |  ||||||||||     |||||    |||
Db    254 ttactgcccagccgtatgaaattgcggacagtaaagaaaatgcaggttatatttccatga 313

Qy    317 ttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaatg 376
          | ||||||||||||| |||||||| |||| || ||  ||||| | ||    || | ||||
Db    314 tcagacaggctgttgagaatgaaactcccgttcatgttttccttaaagtcaacaccaata 373

Qy    377 aaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcctga 436
          |||| | |||||| ||    ||||     || ||| |  | |||||||||| |  |  | |
Db    374 aaattgcaaaagtagaaaaagcaacagatgatgacatccgttatttttaaatctgtattca 433

Qy    437 caaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctgatgtagctacat 496
            ||  |||  | ||  ||| ||| |||    | ||||| |||||||| | || ||
Db    434 acaagcaagagagaggtgaaagcaacaaagcagtcagtgttattcctgatctggcaacgc 493

Qy    497 taaattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcgtcctcac 556
          | ||   |||||| |  || || || || ||| ||||||| || ||||| || || || |
Db    494 tgaacagtttatttacccagatcaaaaaccaggcttgcggaacttctacagcatcttctc 553

Qy    557 catgcatcacattcagatatcctgtagacggatgttatgcaagagcccataagatgagac 616
          | || ||||| |||||||||||| || ||||||||||||||||| || || || |||||||
Db    554 cttgtatcaccttcagatatccggttgacggatgttatgcaagggctcacaaaatgagac 613

Qy    617 aaatcttaatgaacaacggctatgactgtgaaaaacaatttgtatacggaaacctaaagg 676
          ||||| || |||||  |||||||||||||||||| || || |||| || || || |  |
Db    614 aaatcctattgaacgccggctatgactgtgaaaagcagttcgtatatggtaatctgagag 673

Qy    677 catcaacaggaacttgctgtgtggcgtggagctaccacgttgcaatattggtaagctata 736
          | || |||||||| ||||||||  | |||   || ||||| ||||| ||||||||||  |
Db    674 cttctacaggaacatgctgtgtatcatgggtatatcacgtagcaattttggtaagcttca 733

Qy    737 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 796
          |||||||||| ||| |    |||||||||| || ||||||||| |||| || |||||||
Db    734 aaaatgcttcaggaattgttgaaaaaagaatcatagatccttcattattctccagcggtc 793

Qy    797 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 856
          |||||||||| | |||||||||    || ||    |||||    |||||||||| || |
Db    794 ctgtaacagattctgcatggagagctgcatgtaccaacacaagctgcggatctgcgtctg 853

Qy    857 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 916
          | || || || || ||||| |||||||||||||| ||||||||||      ||| |  |
Db    854 tatcttcctacgccaatacagcaggaaatgtttactacagaagtccgtcaggttcattac 913

Qy    917 tgtatgacaacaatctgatcaataccaactgtgtactgactaaattttcactgctttccg 976
          ||||||| |||||     | |||||||| |||||| | |  | ||| |||   ||||| |
Db    914 tgtatgataacaactatgtgaataccaattgtgtattaaacatattctcatccctttcag 973

Qy    977 gatgttctccttcacctgcaccggatgtatccagctgtggattttaatt 1025
          |||||||||||| || |||||   |||| | ||||||||||||||||
Db    974 gatgttctccttccccagcaccaagtgtagcaagctgtggattttaatt 1022
```

RESULT 2
US-09-324-910-5
; Sequence 5, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Chryseobacterium gleum
US-09-324-910-5

Query Match 31.9%; Score 344.2; DB 4; Length 555;
Best Local Similarity 76.7%; Pred. No. 9.1e-91;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy    472 agtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtct 531
          ||||| |||||||||| | || ||  | ||   |||||| |  || || || || ||| ||
Db      7 agtgttattcctgatctggcaacgctgaacagtttatttacccagatcaaaaaccaggct 66

Qy    532 tgcggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgt 591
          ||||| || ||||| || || || || || || ||||| |||||||||||| || |||||||||
Db     67 tgcggaacttctacagcatcttctccttgtatcaccttcagatatccggttgacggatgt 126

Qy    592 tatgcaagagcccataagatgagacaaatcttaatgaacaacggctatgactgtgaaaaa 651
          |||||||| || || || |||||||||||| || |||||  ||||||||||||||||||
Db    127 tatgcaagggctcacaaaatgagacaaatcctattgaacgccggctatgactgtgaaaag 186

Qy    652 caatttgtatacggaaacctaaaggcatcaacaggaacttgctgtgtggcgtggagctac 711
          || || ||||| || || || |  || || |||||||| ||||||||  | |||   ||
Db    187 cagttcgtatatggtaatctgagagcttctacaggaacatgctgtgtatcatgggtatat 246

Qy    712 cacgttgcaatattggtaagctataaaaatgcttccggagtaacggaaaaaagaattatt 771
          ||||| ||||| ||||||||||  |||||||||| ||| |    |||||||||| ||
Db    247 cacgtagcaattttggtaagcttcaaaaatgcttcaggaattgttgaaaaaagaatcata 306

Qy    772 gatccttcactattttcaagcggtcctgtaacagatacagcatggagaaacgcttgcgtt 831
          ||||||||| |||| || |||||||||||||||| | |||||||||   || ||
Db    307 gatccttcattattctccagcggtcctgtaacagattctgcatggagagctgcatgtacc 366

Qy    832 aacacctcttgcggatctgcatccgtttcctcttatgctaatactgcaggaaatgtttat 891
          |||||    |||||||||||| || || || || || || ||||| ||||||||||||||
Db    367 aacacaagctgcggatctgcgtctgtatcttcctacgccaatacagcaggaaatgtttac 426

Qy    892 tacagaagtcctagtaattcttacctgtatgacaacaatctgatcaataccaactgtgta 951
          ||||||||||      ||| |  |||||||| |||||     | ||||||| ||||||
Db    427 tacagaagtccgtcaggttcattactgtatgataacaactatgtgaataccaattgtgta 486

Qy    952 ctgactaaattttcactgctttccggatgttctccttcacctgcaccggatgtatccagc 1011
           | |  | ||| |||   ||||| |||||||||||||| || |||||   |||| | |||
Db    487 ttaaacatattctcatccctttcaggatgttctccttccccagcaccaagtgtagcaagc 546

Qy   1012 tgtggattt 1020
          |||||||||
Db    547 tgtggattt 555
```

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             5.6%;  Score 60;  DB 1;  Length 7218;
  Best Local Similarity   1.6%;  Pred. No. 1.3e-07;
  Matches    6;  Conservative  231;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     126 ttccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagct 185
           ||   ||::::::  ::::::  :: ::::::::::::  ::: : ::: ::    :::
Db    1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy     186 gaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggt 245
           ::::::    :::::::  : :  :: ::::: ::: :::::::: :::: :: :::::
Db    1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy     246 gtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagta 305
           :  :   : :  ::  : : ::    : ::::  :::  ::  :::::::: ::: :: :
Db    1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy     306 tatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcc 365
           :  :::: :   ::: :::  :  ::::: :::      : : : :      ::::
Db    1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy     366 taatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaa 425
           :: :: :: :::: :::::::: ::::   : :::   :::::: : :::: :    ::
Db    1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy     426 aacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctga 485
           :: ::    :: :::::::: ::::::: :::  :: :::  :: ::: : ::     ::
Db    1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1082

Qy     486 tgtagctacattaaattc 503
           : ::  : :  ::: ||
Db    1081 RRRRRRRRRRRRRRATC 1064


RESULT   4
US-09-299-378-3
; Sequence 3, Application US/09299378
; Patent No. 6245545
; GENERAL INFORMATION:
;  APPLICANT: Higgins, Lauren S.
;  APPLICANT: Dalton, Michael A.
;  APPLICANT: Kong, Huimin
;  TITLE OF INVENTION: Method For Cloning And Producing The SwaI Restriction
;  TITLE OF INVENTION: Endonuclease
;  FILE REFERENCE: SwaI
;  CURRENT APPLICATION NUMBER: US/09/299,378
;  CURRENT FILING DATE: 1999-04-27
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1857
;   TYPE: DNA
;   ORGANISM: Staphylococcus warneri
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1854)
US-09-299-378-3


  Query Match             3.9%;  Score 41.8;  DB 4;  Length 1857;
  Best Local Similarity   48.9%;  Pred. No. 0.013;
  Matches  112;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy     302 agtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaa 361
           || |  | |  ||   |  | | ||| |||   ||| |  |  || ||| ||  || |
Db     905 agcaacttgatatcaatttaaaagcttttagtgatgtagttagtggacaaataatcattt 964

Qy     362 agcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatatt 421
             | |   |  |||| |      ||||  |    || ||||    ||||      | ||
Db     965 ctcttttaaaaaatgcatccaacaaagaagctattgtaagtatcaaagattggcaaaata 1024

Qy     422 ttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattc 481
           ||| || || |  |  ||  |||  | | ||       |||||||     || | | |
Db    1025 ataacacaattcaaattaatcaagatatatggttgcatgataaattttatcgtttaata 1084

Qy     482 ctgatgtagctacattaaattctttattcaatcaaataaagaatcagtc 530
           |   |     | | |||||||  | ||  |  || |  | ||| | ||
Db    1085 ttagcgataagaaaataaattcaattttagagaaagtttacaataaatc 1133


RESULT   5
US-09-141-047-7/c
; Sequence 7, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
;  APPLICANT: Walker, David H.
;  APPLICANT: Yu, Xue-Jie
;  TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
;  TITLE OF INVENTION: Protein Gene
;  FILE REFERENCE: D6143
;  CURRENT APPLICATION NUMBER: US/09/141,047A
;  CURRENT FILING DATE: 1998-08-27
;  NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 7
;   LENGTH: 2489
;   TYPE: DNA
;   ORGANISM: Ehrlichia canis
;   FEATURE:
;   LOCATION: -340..2149
;   OTHER INFORMATION: Nucleotide sequence of gene encoding 120kDa
;   OTHER INFORMATION: immunoreactive protein
US-09-141-047-7


  Query Match             3.3%;  Score 35.6;  DB 3;  Length 2489;
  Best Local Similarity   45.9%;  Pred. No. 0.95;
  Matches  122;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy     741 tgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtcctgt 800
           | |||||  | ||     | | |   ||  |   |||||||   |  | |  | | |
Db    1843 TTCTTCCTTACTAGTTTCAGATACTTTTTCTCCAACTTCACTTGATGAATGTTCTACACT 1784

Qy     801 aacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtttc 860
           | ||  |||||||  |  | ||  ||| || |   || ||  | | || | | ||  |
Db    1783 ACCATCTACAGCAGGTTGCAAATCTTCCGCTTTAACTTCAGGAGTATTTTCCTCTTTACT 1724

Qy     861 ctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctgta 920
              ||  | || |  | |   || |     | |   | || |||      | |
Db    1723 AGTTTCAGATACTTTTTCTCCAACTTCACTTGATGAATGTTCTACACTACCATCTACAGC 1664
```

```
Qy    921 tgacaacaatctgatcaataccaactgtgtactgactaaattttcactgctttccggatg 980
          |    |||         |   |  |  | | |    |   | || |||  |||| |
Db   1663 AGGTTGCAAATCTTCTGCTTTAACTTCAGGAGTATTTTCCTCTTTACTAGTTTCAGATAC 1604

Qy    981 ttctccttcacctgcaccggatgtat 1006
          || | || || || |||  |||| ||
Db   1603 TTTTTCTCCAACTTCACTTGATGAAT 1578


RESULT   6
US-09-092-770-8/c
; Sequence 8, Application US/09092770
; Patent No. 5973119
; GENERAL INFORMATION:
;  APPLICANT: Coats, Steven R.
;  APPLICANT: Bass, Michael B.
;  APPLICANT: Robinson, Murray O.
;  TITLE OF INVENTION: No. 5973119el Cyclin E Genes and Proteins
;  FILE REFERENCE: A-524
;  CURRENT APPLICATION NUMBER: US/09/092,770
;  CURRENT FILING DATE: 1997-06-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1215
;   TYPE: DNA
;   ORGANISM: Human
US-09-092-770-8


  Query Match             3.3%;  Score 35.4;  DB 2;  Length 1215;
  Best Local Similarity   24.8%;  Pred. No. 0.77;
  Matches  121;  Conservative   88;  Mismatches  277;  Indels    1;  Gaps    1;

Qy    115 tcctgtgccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaat 174
          ||:|     |  || | : | |:  |  :: | |:|    ::  :|   | |  | : |
Db    773 TCYTTNARNGCRTCNACYTGNARRAANARRTTNARCCANSWDATDATNGTNACNGGRCAN 714

Qy    175 gattctaagctgaaagatttcggaaaga-ctgtaccggtagggatagacgaagaaaacgg 233
          ::||  |  | |: | :||  ::| :|   :: ||    :  ::| :          : :
Db    713 ARYTCCCAYTTNARNGCYTTNARDATDATNARYTCCATNCKNARDATRTCYTCYTCNSWR 654

Qy    234 aatgataaaggtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaaga 293
          | |    :  || |  || |  ||:| :   | : |  |    ||:  :  :   |:
Db    653 CANGCNCCRTCNGTNACRTANGCRAAYTCYTGNARYTTNGGNGCRTADATYTCYTCNARY 594

Qy    294 aaatgagcagtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacat 353
              ||: |: |  :  : | :    |:    ::|   :   : |:|  | :| :  :
Db    593 TTNSWNGCDATRAANARNSWNGTDATNCCDATNARYTGNARCATRTTYTTRTTDATRTCY 534

Qy    354 tttcttaaagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgt 413
          ||:        |  : | | :| :|| | :    | |      |  ||   :    :: |
Db    533 TTYTGNGTNARCATRAANCKRTCRAARAARTCYTGNGCNARRTARAANGTYTCNCKRTGN 474

Qy    414 aagatattttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgag 473
          |:  |:|          |  |:   |: |:        |:   :|  ::    ||: | |
Db    473 ARNGTRTANACYTCRCANACYTCNARNARCCARTCNARNARDATNSWNCKCATYTGNGGY 414

Qy    474 tgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttg 533
          |  |: ||   :|| |:   :   ||:| :||        |: |    :: ||  |:||
Db    413 TCNARRTCNSWRTGNARNACYTCRAARTGYTTRTCRTGNACRTANCKNSWYTCYTTYTTN 354

Qy    534 cggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgtta 593
            :    | :|     :  :    ::|| | |   :: | :|| | |:  |  :   :
Db    353 ARCATRTTNARCCANACYTCYTTNSWRCANCCCCANSWNARRTCNGGNARNGGNSWNGGR 294

Qy    594 tgcaaga 600
          |  | :|
Db    293 TTDATRA 287


RESULT   7
US-09-222-851-8/c
; Sequence 8, Application US/09222851
; Patent No. 6165753
; GENERAL INFORMATION:
;  APPLICANT: Coats, Steven R.
;  APPLICANT: Bass, Michael B.
;  APPLICANT: Robinson, Murray O.
;  TITLE OF INVENTION: No. 6165753el Cyclin E Genes and Proteins
;  FILE REFERENCE: A-524
;  CURRENT APPLICATION NUMBER: US/09/222,851
;  CURRENT FILING DATE: 1998-12-30
;  EARLIER APPLICATION NUMBER: 09/092,770
;  EARLIER FILING DATE: 1998-06-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1215
;   TYPE: DNA
;   ORGANISM: Human
US-09-222-851-8


  Query Match             3.3%;  Score 35.4;  DB 4;  Length 1215;
  Best Local Similarity   24.8%;  Pred. No. 0.77;
  Matches  121;  Conservative   88;  Mismatches  277;  Indels    1;  Gaps    1;

Qy    115 tcctgtgccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaat 174
          ||:|     |  || | : | |:  |  :: | |:|    ::  :|   | |  | : |
Db    773 TCYTTNARNGCRTCNACYTGNARRAANARRTTNARCCANSWDATDATNGTNACNGGRCAN 714

Qy    175 gattctaagctgaaagatttcggaaaga-ctgtaccggtagggatagacgaagaaaacgg 233
          ::||  |  | |: | :||  ::| :|   :: ||    :  ::| :          : :
Db    713 ARYTCCCAYTTNARNGCYTTNARDATDATNARYTCCATNCKNARDATRTCYTCYTCNSWR 654

Qy    234 aatgataaaggtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaaga 293
          | |    :  || |  || |  ||:| :   | : |  |    ||:  :  :   |:
Db    653 CANGCNCCRTCNGTNACRTANGCRAAYTCYTGNARYTTNGGNGCRTADATYTCYTCNARY 594

Qy    294 aaatgagcagtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacat 353
              ||: |: |  :  : | :    |:    ::|   :   : |:|  | :| :  :
Db    593 TTNSWNGCDATRAANARNSWNGTDATNCCDATNARYTGNARCATRTTYTTRTTDATRTCY 534

Qy    354 tttcttaaagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgt 413
          ||:        |  : | | :| :|| | :    | |      |  ||   :    :: |
Db    533 TTYTGNGTNARCATRAANCKRTCRAARAARTCYTGNGCNARRTARAANGTYTCNCKRTGN 474

Qy    414 aagatattttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgag 473
          |:  |:|          |  |:   |: |:        |:   :|  ::    ||: | |
Db    473 ARNGTRTANACYTCRCANACYTCNARNARCCARTCNARNARDATNSWNCKCATYTGNGGY 414

Qy    474 tgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttg 533
          |  |: ||   :|| |:   :   ||:| :||        |: |    :: ||  |:||
Db    413 TCNARRTCNSWRTGNARNACYTCRAARTGYTTRTCRTGNACRTANCKNSWYTCYTTYTTN 354

Qy    534 cggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgtta 593
            :    | :|     :  :    ::|| | |   :: | :|| | |:  |  :   :
Db    353 ARCATRTTNARCCANACYTCYTTNSWRCANCCCCANSWNARRTCNGGNARNGGNSWNGGR 294

Qy    594 tgcaaga 600
          |  | :|
Db    293 TTDATRA 287


RESULT   8
US-09-004-393B-3
; Sequence 3, Application US/09004393B
; Patent No. 6310271
; GENERAL INFORMATION:
;  APPLICANT:  Hanson D., Andrew
;  APPLICANT:  Rathinasabapathi, Bala
;  APPLICANT:  Burnet, Michael
;  TITLE OF INVENTION:  Polynucleotides Encoding Choline Monooxygenase and
;  TITLE OF INVENTION:  Plants Transformed Therewith
;  FILE REFERENCE:  UF-162
```

```
;  CURRENT APPLICATION NUMBER: US/09/004,393B
;  CURRENT FILING DATE:  1998-01-08
;  PRIOR APPLICATION NUMBER:  60/035,147
;  PRIOR FILING DATE:  1997-01-08
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1751
;   TYPE: DNA
;   ORGANISM: Beta vulgaris
US-09-004-393B-3


  Query Match             3.2%;  Score 34.8;  DB 4;  Length 1751;
  Best Local Similarity   48.1%;  Pred. No. 1.4;
  Matches   99;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy    739 aatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtcct 798
          || ||   |  || ||| |||| |  || | | ||||||  |  | ||      |  ||
Db    643 aaagccagcaaagcaactgaaacacaaaatttggatcctaaagaacttgggcttgcaccc 702

Qy    799 gtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccgtt 858
           ||| || | ||| |||| |          ||| || ||    | |||||  |   | |
Db    703 ctaaaagttgcagaatggggcccattcattcttatcagcttggaccgatctctagatgct 762

Qy    859 tcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacctg 918
              || ||| |   || |   | | || | | |    |     |  |||  |  |
Db    763 aatgctgatgttggaacagagtggattggtaaatctgcagaagatgttaaggcccatgct 822

Qy    919 tatgacaacaatctgatcaataccaa 944
          | |||    ||||| |     ||| |
Db    823 tttgatcctaatctaaagttcaccca 848


RESULT   9
US-09-092-770-18/c
; Sequence 18, Application US/09092770
; Patent No. 5973119
; GENERAL INFORMATION:
;  APPLICANT: Coats, Steven R.
;  APPLICANT: Bass, Michael B.
;  APPLICANT: Robinson, Murray O.
;  TITLE OF INVENTION: No. 5973119el Cyclin E Genes and Proteins
;  FILE REFERENCE: A-524
;  CURRENT APPLICATION NUMBER: US/09/092,770
;  CURRENT FILING DATE: 1997-06-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 1212
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-092-770-18


  Query Match             3.2%;  Score 34.6;  DB 2;  Length 1212;
  Best Local Similarity   25.1%;  Pred. No. 1.3;
  Matches  122;  Conservative   85;  Mismatches  279;  Indels    1;  Gaps    1;

Qy    115 tcctgtgccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaat 174
          ||:|   | |  || | : | |:  |  :: | |:|    ::   |   | |  | : |
Db    770 TCYTTNACNGCRTCNACYTGNARRAANARRTTNARCCANSWDATNACNGTNACNGGRCAN 711

Qy    175 gattctaagctgaaagatttcggaa-agactgtaccggtagggatagacgaagaaaacgg 233
           ::||  |  | |: | :||  ::| :    :: ||  :    ::| :   |     : :
Db    710 ARYTCCCAYTTNARNGCYTTNARDATRTTNARYTCCATYTTNARDATRTCNACYTCNSWR 651

Qy    234 aatgataaaggtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaaga 293
          | |    :  || |  || |  ||:| :   | : |  |    ||:  :  :   |:
Db    650 CANGCNCCRTCNGTNACRTANGCRAAYTCYTGNARYTTNGGNGCRTADATYTCYTCNARY 591

Qy    294 aaatgagcagtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacat 353
                ||: |: |  :  : | :    |:    ::|   :   : |:|  |  ||:  :
Db    590 TTNSWNGCDATRAANARNSWNGTDATNCCDATNARYTGNARCATRTTYTTRTTNACRTCY 531

Qy    354 tttcttaaagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgt 413
          ||:        |  : | | :| :|| | :    | |      |  ||   :    :: |
Db    530 TTYTGNGTNARCATRAANCKRTCRAARAARTCYTGNGCNARRTARAANGTYTCNCKRTGN 471

Qy    414 aagatattttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgag 473
          |:  |:|          |  |:   |: |:       |:   :|  ::    ||: | |
Db    470 ARNGTRTANACYTCRCANACYTCNARNARCCARTCNARNARDATNSWNCKCATYTGNGGY 411

Qy    474 tgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttg 533
          |  |: ||   :|| |:   :   ||:| :||        |: |   :   ||  |:|
Db    410 TCNARRTCNSWRTGNARNACYTGRAARTGYTTRTCRTGNACRTANCKRTTYTCYTTYTGN 351

Qy    534 cggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgtta 593
           :    |:| |     : :   ::||   |   :: | :||  | |:  |  :   :
Db    350 ARCATRTTYTGCCANACYTCYTGNSWRCANGCCCANSWNARRTCNGGNARNGGNSWNGGR 291

Qy    594 tgcaaga 600
          |  | :|
Db    290 TTDATRA 284


RESULT  10
US-09-222-851-18/c
; Sequence 18, Application US/09222851
; Patent No. 6165753
; GENERAL INFORMATION:
;  APPLICANT: Coats, Steven R.
;  APPLICANT: Bass, Michael B.
;  APPLICANT: Robinson, Murray O.
;  TITLE OF INVENTION: No. 6165753el Cyclin E Genes and Proteins
;  FILE REFERENCE: A-524
;  CURRENT APPLICATION NUMBER: US/09/222,851
;  CURRENT FILING DATE: 1998-12-30
;  EARLIER APPLICATION NUMBER: 09/092,770
;  EARLIER FILING DATE: 1998-06-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 1212
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-222-851-18


  Query Match             3.2%;  Score 34.6;  DB 4;  Length 1212;
  Best Local Similarity   25.1%;  Pred. No. 1.3;
  Matches  122;  Conservative   85;  Mismatches  279;  Indels    1;  Gaps    1;

Qy    115 tcctgtgccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaat 174
          ||:|   | |  || | : | |:  |  :: | |:|    ::   |   | |  | : |
Db    770 TCYTTNACNGCRTCNACYTGNARRAANARRTTNARCCANSWDATNACNGTNACNGGRCAN 711

Qy    175 gattctaagctgaaagatttcggaa-agactgtaccggtagggatagacgaagaaaacgg 233
           ::||  |  | |: | :||  ::| :    :: ||  :    ::| :   |     : :
Db    710 ARYTCCCAYTTNARNGCYTTNARDATRTTNARYTCCATYTTNARDATRTCNACYTCNSWR 651

Qy    234 aatgataaaggtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaaga 293
          | |    :  || |  || |  ||:| :   | : |  |    ||:  :  :   |:
Db    650 CANGCNCCRTCNGTNACRTANGCRAAYTCYTGNARYTTNGGNGCRTADATYTCYTCNARY 591

Qy    294 aaatgagcagtatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacat 353
                ||: |: |  :  : | :    |:    ::|   :   : |:|  |  ||:  :
Db    590 TTNSWNGCDATRAANARNSWNGTDATNCCDATNARYTGNARCATRTTYTTRTTNACRTCY 531

Qy    354 tttcttaaagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgt 413
          ||:        |  : | | :| :|| | :    | |      |  ||   :    :: |
Db    530 TTYTGNGTNARCATRAANCKRTCRAARAARTCYTGNGCNARRTARAANGTYTCNCKRTGN 471

Qy    414 aagatattttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgag 473
          |:  |:|          |  |:   |: |:       |:   :|  ::    ||: | |
Db    470 ARNGTRTANACYTCRCANACYTCNARNARCCARTCNARNARDATNSWNCKCATYTGNGGY 411
```

```
Qy    474 tgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttg 533
Db    410 TCNARRTCNSWRTGNARNACYTGRAARTGYTTRTCRTGNACRTANCKRTTYTCYTTYTGN 351

Qy    534 cggtacctctacggcgtcctcaccatgcatcacattcagatatcctgtagacggatgtta 593
Db    350 ARCATRTTYTGCCANACYTCYTGNSWRCANGCCCANSWNARRTCNGGNARNGGNSWNGGR 291

Qy    594 tgcaaga 600
Db    290 TTDATRA 284


RESULT  11
US-08-417-492-1
; Sequence 1, Application US/08417492
; Patent No. 5750872
;  GENERAL INFORMATION:
;    APPLICANT:  Bennett, Alan B
;    APPLICANT:  Brummell, David A
;    APPLICANT:  Grantz, Alexander A
;    TITLE OF INVENTION:  Nucleic Acids Encoding Ascorbate Free
;    TITLE OF INVENTION:  Radical Reductase and Their Uses
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Street Tower
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105-1492
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/417,492
;      FILING DATE:  05-APR-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin L
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  2307E-586US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-543-9600
;      TELEFAX:  415-543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  49..643
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  644..780
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  781..1484
;      OTHER INFORMATION:  /number= 2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(49..643, 781..1487)
US-08-417-492-1
```

```
  Query Match             3.2%;  Score 34.2;  DB 1;  Length 1723;
  Best Local Similarity   46.4%;  Pred. No. 2;
  Matches  111;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy    159 aaaactaagtgtaaatgattctaagctgaaagatttcggaaagactgtaccggtagggat 218
Db    890 AAATGGAGAGGTGAAGGAAGTCAAACTCAAAGATGGCAGAGTTTTGGAAGCTGACATAGT 949

Qy    219 agacgaagaaaaacggaatgataaaggtgtcatttatgttaactgcgcaattctatgaaat 278
Db    950 AGTCGTAGGAGTCGGAGCAAGACCACTCACAACTCTATTCAAAGGGCAAGTTGAAGAGGA 1009

Qy    279 taagccgaccaaagaaaatgagcagtatatcggaatgcttagacaggctgttaagaatga 338
Db   1010 GAAGGGTGGAATTAAGACAGATGCGTTCTTCAAAACAAGTGTACCTGATGTATATGCTGT 1069

Qy    339 atctcctgtacacattttcttaaagcctaatagcaatgaaataggaaaagtggagtctg 397
Db   1070 GGGTGATGTTGCCACTTTTCCTTTGAAAATGTACAATGAGATTAGAAGAGTTGAACATG 1128


RESULT  12
US-08-682-517-13
; Sequence 13, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3666 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-682-517-13


  Query Match             3.2%;  Score 34.2;  DB 2;  Length 3666;
  Best Local Similarity   47.4%;  Pred. No. 2.9;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    737 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 796
Db   1013 AAGGTACTGATGAAGTAACTGCTTATGCAACTGGTGATCGTTCTAAATTCTCACTTGGTT 1072

Qy    797 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 856
Db   1073 ATGTATTCTGGGGTGTAGATACAATTCTTTCAGTTGAAGAAGTAACTACAGGTGCTTCAG 1132

Qy    857 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 916
Db   1133 TTAATAATGGTGCAAACAAAACTTACAAAGTTACTTATAAAAACCCTAAAACTGGTAAAC 1192

Qy    917 tgtatgacaacaatctgatcaataccaactgtgta 951
Db   1193 CAGAAGCAAACAAAACATTTAATGTTGGTTTTGTA 1227


RESULT  13
US-08-682-517-14
; Sequence 14, Application US/08682517
; Patent No. 5874267
```

```
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3666 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3666
US-08-682-517-14


  Query Match             3.2%;  Score 34.2;  DB 2;  Length 3666;
  Best Local Similarity   47.4%;  Pred. No. 2.9;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    737 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 796
          ||  | ||   | |||||| |   |   || | ||||| |||   ||| |||   |||
Db   1013 AAGGTACTGATGAAGTAACTGCTTATGCAACTGGTGATCGTTCTAAATTCTCACTTGGTT 1072

Qy    797 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 856
           ||||         | |   | ||    ||  ||| |           |  ||| || |
Db   1073 ATGTATTCTGGGGTGTAGATACAATTCTTTCAGTTGAAGAAGTAACTACAGGTGCTTCAG 1132

Qy    857 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 916
          ||     |  ||| || |   |    || |||  ||| | ||  |||| | |  | | |
Db   1133 TTAATAATGGTGCAAACAAAACTTACAAAGTTACTTATAAAAACCCTAAAACTGGTAAAC 1192

Qy    917 tgtatgacaacaatctgatcaataccaactgtgta 951
            | |  |||||     | |||      | ||||
Db   1193 CAGAAGCAAACAAAACATTTAATGTTGGTTTTGTA 1227


RESULT  14
US-08-682-517-7
; Sequence 7, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-682-517-7


  Query Match             3.2%;  Score 34.2;  DB 2;  Length 4197;
  Best Local Similarity   47.4%;  Pred. No. 3.1;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    737 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 796
          ||  | ||   | |||||| |   |   || | ||||| |||   ||| |||   |||
Db   1197 AAGGTACTGATGAAGTAACTGCTTATGCAACTGGTGATCGTTCTAAATTCTCACTTGGTT 1256

Qy    797 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 856
           ||||         | |   | ||    ||  ||| |           |  ||| || |
Db   1257 ATGTATTCTGGGGTGTAGATACAATTCTTTCAGTTGAAGAAGTAACTACAGGTGCTTCAG 1316

Qy    857 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 916
          ||     |  ||| || |   |    || |||  ||| | ||  |||| | |  | | |
Db   1317 TTAATAATGGTGCAAACAAAACTTACAAAGTTACTTATAAAAACCCTAAAACTGGTAAAC 1376

Qy    917 tgtatgacaacaatctgatcaataccaactgtgta 951
            | |  |||||     | |||      | ||||
Db   1377 CAGAAGCAAACAAAACATTTAATGTTGGTTTTGTA 1411


RESULT  15
US-08-682-517-8
; Sequence 8, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacillus sphaericus
;      INDIVIDUAL ISOLATE:  P-1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  95..3850
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  185..3850
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  95..184
US-08-682-517-8


  Query Match             3.2%;  Score 34.2;  DB 2;  Length 4197;
  Best Local Similarity   47.4%;  Pred. No. 3.1;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    737 aaaatgcttccggagtaacggaaaaaagaattattgatccttcactattttcaagcggtc 796
          ||  | ||   | |||||| |   |   || | ||||| |||   ||| |||   |||
Db   1197 AAGGTACTGATGAAGTAACTGCTTATGCAACTGGTGATCGTTCTAAATTCTCACTTGGTT 1256

Qy    797 ctgtaacagatacagcatggagaaacgcttgcgttaacacctcttgcggatctgcatccg 856
           ||||         | |   | ||    ||  ||| |           |  ||| || |
Db   1257 ATGTATTCTGGGGTGTAGATACAATTCTTTCAGTTGAAGAAGTAACTACAGGTGCTTCAG 1316
```

```
Qy     857 tttcctcttatgctaatactgcaggaaatgtttattacagaagtcctagtaattcttacc 916
           ||     |  ||| || |   |    || |||  ||| | ||  ||||  | |  | | |
Db    1317 TTAATAATGGTGCAAACAAAACTTACAAAGTTACTTATAAAAACCCTAAAACTGGTAAAC 1376

Qy     917 tgtatgacaacaatctgatcaataccaactgtgta 951
             | |  |||||     | |||      | ||||
Db    1377 CAGAAGCAAACAAAACATTTAATGTTGGTTTGTA 1411


Search completed: June 27, 2002, 20:15:27
Job time: 9570 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 19:22:05 ; Search time 2534.16 Seconds
(without alignments)
5752.088 Million cell updates/sec

Title: US-09-727-769A-7
Perfect score: 1080
Sequence: 1 agttaaaataaccaaccaac.....atcagcaggtgctgttatat 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	55.8	5.2	1101	12	CNS0039G	AL063921 Drosophil
	2	51.2	4.7	1101	12	CNS0182P	AL108811 Drosophil
	3	50.2	4.6	893	12	AZ538298	ENTGB51TR
	4	50.2	4.6	918	12	BH162500	BH162500 ENTRU42TR
c	5	50	4.6	942	12	CNS018GS	AL109318 Drosophil
c	6	48	4.4	769	12	CNS011WN	AL100817 Drosophil
	7	47.8	4.4	734	12	CNS010MP	AL099163 Drosophil
	8	46.6	4.3	878	12	CNS0187R	AL108993 Drosophil
	9	46	4.3	1007	12	CNS06X9S	AL419462 T3 end of
	10	46	4.3	1101	12	CNS00LJT	AL068307 Drosophil
c	11	45.8	4.2	997	12	CNS005TE	AL060767 Drosophil
c	12	45.6	4.2	816	12	CNS010PW	AL099278 Drosophil
	13	45.6	4.2	965	12	BH137450	BH137450 ENTOI31TF
	14	45	4.2	581	12	CNS01UPZ	AL168128 Tetraodon
c	15	45	4.2	964	12	CNS07E8R	AL441457 T7 end of
	16	45	4.2	1118	12	CNS02PY4	AL208597 Tetraodon
	17	44.4	4.1	781	12	CNS009DO	AL053444 Drosophil
c	18	44.4	4.1	928	12	CNS00DKY	AL071865 Drosophil
c	19	44.2	4.1	773	12	BH534872	BH534872 BOHAG04TF
	20	44.2	4.1	959	12	CNS00655	AL062806 Drosophil
	21	43.8	4.1	1101	12	CNS016LI	AL106896 Drosophil
c	22	43.6	4.0	1101	12	CNS00HX9	AL073856 Drosophil
c	23	43.2	4.0	1101	12	CNS00Z15	AL097091 Drosophil
	24	43	4.0	1190	12	CNS02ON7	AL206908 Tetraodon
	25	42.8	4.0	1101	12	CNS003B4	AL064082 Drosophil
c	26	42.8	4.0	1101	12	CNS00LT2	AL078714 Drosophil
	27	42.6	3.9	1101	12	CNS00EVL	AL069706 Drosophil
	28	42.4	3.9	886	12	CNS075DF	AL429961 clone BA0
c	29	42.4	3.9	930	12	AZ693378	AZ693378 ENTIE73TR
c	30	42.4	3.9	969	12	BH154338	BH154338 ENTQY44TF
c	31	42.4	3.9	988	12	CNS00ZHB	AL097673 Drosophil
c	32	42.4	3.9	1204	12	CNS016E2	AL106628 Drosophil
	33	42.2	3.9	562	12	CNS016OR	AL107013 Drosophil
c	34	42.2	3.9	974	12	CNS00ITT	AL075432 Drosophil
c	35	42.2	3.9	991	12	CNS00I2D	AL074343 Drosophil
	36	42.2	3.9	1101	12	CNS012JN	AL101645 Drosophil
	37	42	3.9	760	12	CNS011DI	AL100128 Drosophil
c	38	42	3.9	827	12	CNS00EU3	AL069854 Drosophil
	39	42	3.9	911	12	BH163485	BH163485 ENTSH90TF
	40	41.8	3.9	617	12	CNS031FC	AL223473 Tetraodon
	41	41.8	3.9	888	12	AZ547775	AZ547775 ENTFG64TR
c	42	41.8	3.9	938	12	CNS006TJ	AL065906 Drosophil
c	43	41.8	3.9	1201	12	CNS0167M	AL106396 Drosophil
c	44	41.6	3.9	963	9	AL566565	AL566565 AL566565
	45	41.6	3.9	1101	12	CNS017V6	AL108540 Drosophil

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"

```
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN


  Query Match             5.2%;  Score 55.8;  DB 12;  Length 1101;
  Best Local Similarity   14.6%;  Pred. No. 0.011;
  Matches   71;  Conservative  228;  Mismatches  186;  Indels    0;  Gaps    0;

Qy      48 aattacaattattatgaaaaatctttttttatcaatgatggcctttgtgaccgtcttaac 107
           :::::  ::|::::  :::::  : : : ::: | | :|    ::: : |  :| :: :
Db    1084 DDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWWWWATWDTWWDKWW 1025

Qy     108 ttttaattcctgtgccgattccaacgggaatcaggaaatcaacggaaaggaaaaactaag 167
           :: |||:|    |    |:   :| ::: |  :::|||  |:  |:::||: : : ::::
Db    1024 WWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKRKDKKDRK 965

Qy     168 tgtaaatgattctaagctgaaagatttcggaaagactgtaccggtagggatagacgaaga 227
           :|::: :| :: :||:  ::| ::::: : : :   :|    : :::::|| ::: :  |
Db     964 DGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDGKGK 905

Qy     228 aaacggaatgataaaggtgtcatttatgttaactgcgcaattctatgaaattaagccgac 287
           :|: : : |::::: ::::: :: : |  :|       ::::  : :|:::|::   ::
Db     904 DADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDW 845

Qy     288 caaagaaaatgagcagtatatcggaatgcttagacaggctgttaagaatgaatctcctgt 347
           |:::::|:     : :|:     :: :  ::   :||  :  ::::|:: ::     ::
Db     844 WADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDD 785

Qy     348 acacattttcttaaagcctaatagcaatgaaataggaaaagtggagtctgcaagtccgga 407
           |    |:|| ||:: :  :: |:   :: |::|:::::::: ::|||  : :::|  :::
Db     784 AATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRW 725

Qy     408 agacgtaagatattttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaatt 467
           ::   :::|:|::| :::: ::     :     :  :::::::  :::  |::|:| :
Db     724 KRRDTRWDDADADDTARDDRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAW 665

Qy     468 ggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatca 527
           : : :::::   :::| ::  :    ||::: |::|:: | : :   |:: ::: |
Db     664 WTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAA 605

Qy     528 gtctt 532
           :| |:
Db     604 DTWTD 600


RESULT   2
CNS0182P
LOCUS       CNS0182P                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108811
VERSION     AL108811.1  GI:5629115
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37D10"
                     /note="end : SP6"
BASE COUNT      274 a    268 c    128 g     73 t    358 others
ORIGIN


  Query Match             4.7%;  Score 51.2;  DB 12;  Length 1101;
  Best Local Similarity   19.4%;  Pred. No. 0.13;
  Matches   72;  Conservative  150;  Mismatches  150;  Indels    0;  Gaps    0;

Qy      23 aacaaaaactcaccattaaactacaaattacaattattatgaaaaatctttttttatcaa 82
           |: |:|||   |  |  |||  | |||  | |:   :  ::  :  :  : |:| |:
Db     683 ARAARAAAAADAKRAAAAAAAAAAAAAAAAAAAADAGGKRKWKKGDGKDGGKATKTAAWAKG 742

Qy      83 tgatggcctttgtgaccgtcttaacttttaattcctgtgccgattccaacgggaatcagg 142
           :::|    :: ::|   : ::|| :|:  :     ::  : |       : : :  |
Db     743 RKDWGTATAWWTWDATWADTWKAATDTDAKRAAAGRRKRDARKTARDGGRRARTRRRAWA 802

Qy     143 aaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagatttcggaaaga 202
           |  : |: | ::|: |::|| ::: : :||  |    ||  | :::::::: : : : :
Db     803 AGGKRARAGARRARAARRAADDRDWDAWAAAAAAAAAAAAAAAWTTWRDRWWDWDDWDTRWD 862

Qy     203 ctgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatttatgttaactg 262
           : ||   ::|: :| ::: ::::::: : |:  |:: : :::: : ||:::||    |
Db     863 DDTTAAWWDDARARARRRRRRRRRRRRRARRAARRAADDTDTKDRWADATTDKDTTKTWTT 922

Qy     263 cgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaatgcttagac 322
           : :::: ::::::| :|:|  ::  :|::::|||  : :|: : : |:   | :: ::
Db     923 DDDDWDKAKRDRWWAAKADGAWKWRDRARDDWAATAKDDDGWKDKWGGRGKRGGKDKKRW 982

Qy     323 aggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaatgaaatag 382
           ::| ||:::::::: :| :  : :    ::::: : : :|  |:  :| ::  :|: ::
Db     983 DKGKTGKKDDDDWDKTWTRDWWWWWTRTKWDWWWDDGRGRGRWTRKRWGAWWRADAWARD 1042

Qy     383 gaaaagtggagt 394
           : :  :|: |::
Db    1043 DTDGKDTRTADK 1054


RESULT   3
AZ538298
LOCUS       AZ538298                893 bp    DNA     linear   GSS 14-NOV-2000
DEFINITION  ENTGB51TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ538298
VERSION     AZ538298.1  GI:11143106
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
```

```
            Class: shotgun
            High quality sequence start: 25
            High quality sequence stop: 829.
FEATURES             Location/Qualifiers
     source          1. .893
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      364 a     98 c     87 g    344 t
ORIGIN


  Query Match             4.6%;  Score 50.2;  DB 12;  Length 893;
  Best Local Similarity   48.7%;  Pred. No. 0.21;
  Matches  136;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy      3 ttaaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattat 62
          ||| |||||||||  |  | |  | |   | |  |  | |  ||||||  | |   | ||
Db    352 TTATAATAACCAATTATATAATAATATTTTAATTAAAAGAAAACAAATAATAGAAACTAC 411

Qy     63 gaaaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgc 122
            | || ||  ||||||||||||  |  | | |  |||  ||   |  || ||     |
Db    412 TTCATATATTGATTTATCAATGAAAGATTCTATTGCCGAGTTTTATGATAGTTTAATGGA 471

Qy    123 cgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaa 182
           |  |   |   | |||   ||  |    |   ||| || |   | | |||||| |||
Db    472 AGCATATCATTCAATTCAAACAACTATTCAAGTAGAACAAGTTGTTTTTAATGATTATAA 531

Qy    183 gctgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaa 242
            |   ||| | ||| |  | | | || ||     ||  |     ||   | ||
Db    532 TGTTGCACCATTAGTAAAACATAAATCTATTGGAATTACTAAAACACTTCCAAAATTTAA 591

Qy    243 ggtgtcatttatgttaactgcgcaattctatgaaattaa 281
           |    | || | || | | |   |   ||| ||| | |
Db    592 TGAACGAATTTTATTTAATACAACAAAATATAAAAATGA 630


RESULT   4
BH162500
LOCUS       BH162500                 918 bp    DNA     linear   GSS 24-SEP-2001
DEFINITION  ENTRU42TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH162500
VERSION     BH162500.1  GI:15735938
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 710.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      385 a    105 c     92 g    336 t
ORIGIN


  Query Match             4.6%;  Score 50.2;  DB 12;  Length 918;
  Best Local Similarity   48.7%;  Pred. No. 0.21;
  Matches  136;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy      3 ttaaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattat 62
          ||| |||||||||  |  | |  | |   | |  |  | |  ||||||  | |   | ||
Db    167 TTATAATAACCAATTATATAATAATATTTTAATTAAAAGAAAACAAATAATAGAAACTAC 226

Qy     63 gaaaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgc 122
            | || ||  ||||||||||||  |  | | |  |||  ||   |  || ||     |
Db    227 TTCATATATTGATTTATCAATGAAAGATTCTATTGCCGAGTTTTATGATAGTTTAATGGA 286

Qy    123 cgattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaa 182
           |  |   |   | |||   ||  |    |   ||| || |   | | |||||| |||
Db    287 AGCATATCATTCAATTCAAACAACTATTCAAGTAGAACAAGTTGTTTTTAATGATTATAA 346

Qy    183 gctgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaa 242
            |   ||| | ||| |  | | | || ||     ||  |     ||   | ||
Db    347 TGTTGCACCATTAGTAAAACATAAATCTATTGGAATTACTAAAACACTTCCAAAATTTAA 406

Qy    243 ggtgtcatttatgttaactgcgcaattctatgaaattaa 281
           |    | || | || | | |   |   ||| ||| | |
Db    407 TGAACGAATTTTATTTAATACAACAAAATATAAAAATGA 445


RESULT   5
CNS018GS/c
LOCUS       CNS018GS                 942 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN13P09 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL109318
VERSION     AL109318.1  GI:5629622
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 942)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
```

```
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..942
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN13P09"
                     /note="end : T7"
BASE COUNT      155 a     77 c     70 g    494 t    146 others
ORIGIN


  Query Match             4.6%;  Score 50;  DB 12;  Length 942;
  Best Local Similarity   34.0%;  Pred. No. 0.24;
  Matches  179;  Conservative   60;  Mismatches  288;  Indels    0;  Gaps    0;

Qy    130 aacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaa 189
           ||  |||  | | ||  ||  ||||:||| ||  ||       |||  |    ||   |||
Db    603 AAAAMGAAAAAAGCAAAAAAACGAAARGAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 544

Qy    190 gatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtca 249
            |    | ||| |   |    |          | :  || ||||  |  |  |::::     ||
Db    543 AAAAAAGAAAAAAAATTNGAGAACNAACGAAWAAAAAAAAAAACGTAAAAWWWWWWAAACA 484

Qy    250 tttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatc 309
               : :  ||       |        |||  ||    |  ||| |:| |:| ::: |
Db    483 AGARARAAAAANANNNNANNANNNAAAANAAAAAAAAAAAAAAAMACTRAAMWRAAATAA 424

Qy    310 ggaatgcttagacaggctgttaagaatgaatctcctgtacacatttcttaaagcctaat 369
           : ||   : | |:|          || ||: || |   | |  :   :  |||      |||
Db    423 RAAAAAAWAAAAMACMRAARAAAAAAWWAACCGGGTARAWWTTWAAWTGTAACCRAAAAT 364

Qy    370 agcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacg 429
            |  ||  ||| |  ||||        |:|    ||        || |   || | |   |||:
Db    363 AAAAAAAAAAAAAAAAAAAWAAWARTAAMAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWAA 304

Qy    430 atcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctgatgta 489
            :     :    :| :|  |::|     |||  ||:|||         |    :||  :: :: :
Db    303 WAWWATMACCWAAMAAAAWWAAAAAAAAAAAAWAAAAAAAWAAAAAWAAACMMAAWWWMW 244

Qy    490 gctacattaaattctttattcaatcaaataaagaatcagtcttgcggtacctctacggcg 549
               : | :|:| : || :|  ||  ||:: | |:  : |:       :|          :|
Db    243 AAAWAAAWAWAAWWTTAWTAAAAAWAAWWTGAWAWAAWATMGGTTAAWAAAATCGTTAMG 184

Qy    550 tcctcaccatgcatcacattcagatatcctgtagacggatgttatgcaagagcccataag 609
                |  ::  |: | |       : |   :| | |  |||    |  || :      |:||
Db    183 AAAAWAAAWWAAAWWAWAAAATAWAAAWAWGCAAAATTATGAATTWAAAAWAAWAAWAAW 124

Qy    610 atgagacaaatcttaatgaacaacggctatgactgtgaaaaacaatt 656
           :: | | :||:   | | |:  || ||       |  :  ||  |||
Db    123 WWWAAAAWAAWAAAAGTAAWWAAAWGCGATTTGRTTTRGTAAWCATT 77


RESULT     6
CNS011WN/c
LOCUS       CNS011WN             769 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN07M08 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL100817
VERSION     AL100817.1  GI:5612428
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 769)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..769
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN07M08"
                     /note="end : T7"
BASE COUNT       58 a     22 c     47 g    381 t    261 others
ORIGIN


  Query Match             4.4%;  Score 48;  DB 12;  Length 769;
  Best Local Similarity   28.1%;  Pred. No. 0.69;
  Matches  144;  Conservative   17;  Mismatches  352;  Indels    0;  Gaps    0;

Qy     12 ccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatgaaaaatct 71
          ::::::| :|||: |::  :::|:|: : ||  | | |  | ||  |      ||||
Db    659 YYMWMYAYMTTAWAAMMCKYKYAYCWCYTAAAAAAACAAAAAAAGAAAANAAAAAATANN 600

Qy     72 tttttatcaatgatggcctttgtgaccgtcttaactttaattcctgtgccgattccaa 131
              |  |         |   | |          |                  |    ||
Db    599 NCAAGAANAANNNNNNANNTNNNNAAACTANNNNNANAATNNNNNANNNNNNNAAAAAAA 540

Qy    132 cgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaaga 191
             || | |||  ||   |||  |||||  |      |        |      ||| |
Db    539 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNANNANAANNNNNANNNNNAAANAAAAANA 480

Qy    192 tttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatt 251
             | |      |          | | |       |||| |               |
Db    479 NNNNNNANAANNNNNAANNNNANNANAAACAAANTAAAAANGANNANGNNNNATANAANTA 420

Qy    252 tatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcgg 311
                        ||         |  ||    |  ||| ||||  |   |   |
Db    419 NNANNANNANAANAAANNANNNTNNAANAAAAAAAAAAAAAAAAAANAAAAAAAAANCNNN 360

Qy    312 aatgcttagacaggctgttaagaatgaatctcctgtacacatttcttaaagcctaatag 371
                   |             | ||  |                       | |    |||
Db    359 NNNAACNNNANNANNAANNGAAAANAANNNANANNANANNNANANNAAAGAAANNNATAA 300

Qy    372 caatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgat 431
            | ||||  |      || ||        | | ||                |     |      |
Db    299 AAGGGAAANANAGNNAGNNGACAAACACTTTNAGTAANTNANTTANNANNCNNANNANAA 240

Qy    432 cctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcctgatgtagc 491
              | | |       ||||     ||  || ||||          | :         | | |   ||
Db    239 NNAAAAANANNNNNCAAAAAAANNTAAAAAAAAAATNNNNNAAWACCNANCNTNAAATANN 180

Qy    492 tacattaaattctttattcaatcaaataaagaa 524
            |     ||       | ||||  ||| ||| ||
Db    179 CAANNCAACNNANAAAATCAAAAAAAAAAAAA 147


RESULT     7
CNS010MP
```

```
LOCUS       CNS010MP                734 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099163
VERSION     AL099163.1  GI:5610774
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN04L20"
                     /note="end : T7"
BASE COUNT      288 a     62 c      2 g    211 t    171 others
ORIGIN
```

```
  Query Match             4.4%;  Score 47.8;  DB 12;  Length 734;
  Best Local Similarity   34.7%;  Pred. No. 0.76;
  Matches  124;  Conservative   54;  Mismatches  179;  Indels    0;  Gaps    0;

Qy      4 taaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatg 63
          ||:|| |: ::  ||| || ::|||| : |  ::||::: : | |||:   ||  | |
Db     79 TAWAAAAMAMWTTCAAATTTMMAAAAAAWWAAAWWTAWHMAWAATATTWATTTTTMTTTT 138

Qy     64 aaaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgcc 123
          ::||:   ||||||     | |     ||| | |   :: | |  ||| :|  |    :
Db    139 HMAAWATATTTTTTTMATTTAAAAATTTTTTTAAAWWWMMTYAWATTTTTWTWATWAAWM 198

Qy    124 gattccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaag 183
           :   |:   |:  |  | || |:   ||:  |:|:| ||  :  |||  ::: ::
Db    199 WTWAAAAWAAWAAWATTATATTAWTTTAAMAAAWAMAATATWWMAAAAAMWWWWWTA 258

Qy    184 ctgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaag 243
            |:| | ::|   || |     |     :    ||    ::  |:|   :|  | |||
Db    259 TAMAWAAAAWWCAACAAAAAATMAAAAWCWATATTAATTTWWMCAHAAAAWAATACAAAA 318

Qy    244 gtgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcag 303
           | |:| ::|    |:      ||:: :   | |: ||    : || | ||:|  |  |
Db    319 WTATMAAWWAAAAAAWAAATAWAAWWAWTAAATAWAAAAWAAMACATAAAAWAMTATTAC 378

Qy    304 tatatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttctta 360
            |:||   |:  |  |||   |   ||  |  ||| |  |     |  | ||  ||
Db    379 AAWATWMATAWWTTATAACAACATAAAAAWCACAAATATTWTTATTAWTTCTTTATA 435
```

```
RESULT   8
CNS0187R
LOCUS       CNS0187R                878 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN04E04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108993
VERSION     AL108993.1  GI:5629297
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .878
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN04E04"
                     /note="end : SP6"
BASE COUNT      279 a    132 c    120 g    182 t    165 others
ORIGIN
```

```
  Query Match             4.3%;  Score 46.6;  DB 12;  Length 878;
  Best Local Similarity   25.2%;  Pred. No. 1.5;
  Matches   76;  Conservative   88;  Mismatches  137;  Indels    0;  Gaps    0;

Qy    227 aaaacggaatgataaaggtgtcatttatgttaactgcgcaattctatgaaattaagccga 286
          ||||   ||| | |::  ||  |   |: :  |      ||  :::  |||: :|  : :
Db     41 AAAAAWMAATCAAAWRTATGAWAAAAAWMWMCAAAAAAAAAAAMMWMAAAAAWAWAAAMAW 100

Qy    287 ccaaagaaaatgagcagtatatcggaatgcttagacaggctgttaagaatgaatctcctg 346
          : :|| ||||: : :|: |:|:   :    : : | :| :  : :: ||  |:::: ||
Db    101 MWMAAAAAAAWWWAMARGAWAWGCWWGGAWWMMMAWRGMMMWAWWAAAAAAWWMWWCTT 160

Qy    347 tacacattttcttaaagcctaatagcaatgaaataggaaaagtggagtctgcaagtccgg 406
          | |:|   :| |:   :  |:|:: :: |:||:::  |:::    : :  | ::   :
Db    161 TGCMCTGAWTRTWGTTRTWTWAWWWMRTTRAAWWWMAAWWWMAAWWWWARGWWWWRGMMA 220

Qy    407 aagacgtaagatattttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaat 466
          || :  ::| :::   : : :| | ::  : |||| ||  ||:|:   ::: :|| : :
Db    221 AAARAWWWAMMWRCRCWTWCRCMAAMMAAMAAAAAAMAAMMAAMARCAWMWMWMAAMWCRA 280

Qy    467 tggcgagtgtaattcctgatgtagctacattaaattctttattcaatcaaataaagaatc 526
             ||  : |:|   : : : :   |::  ::|:  :  |   |:   :   |: ||:
Db    281 AWMGGAWARAAWTRGAWWWMRAMMAAAMRAARMAWRAWMMAAAWAMAATWTACAWAAAWA 340

Qy    527 a 527
          |
Db    341 A 341
```

```
RESULT   9
CNS06X9S
LOCUS       CNS06X9S                1007 bp    DNA     linear   GSS 06-JUL-2001
DEFINITION  T3 end of clone AX0AA039F08 of library AX0AA from strain CBS 7064
            of Pichia farinosa, genomic survey sequence.
ACCESSION   AL419462
VERSION     AL419462.1  GI:12202640
KEYWORDS    GSS.
SOURCE      Pichia farinosa.
  ORGANISM  Pichia farinosa
```

```
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 1007)
  AUTHORS   de Montigny,J., Spehner,C., Souciet,J., Tekaia,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Potier,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
            sorbitophila
  JOURNAL   FEBS Lett. 487 (1), 87-90 (2000)
  MEDLINE   20584725
REFERENCE   2  (bases 1 to 1007)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
REFERENCE   3  (bases 1 to 1007)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1007
                     /organism="Pichia farinosa"
                     /strain="CBS 7064"
                     /db_xref="taxon:4920"
                     /clone="AX0AA039F08"
                     /clone_lib="AX0AA"
                     /note="end : T3"
BASE COUNT      533 a     86 c    126 g    161 t    101 others
ORIGIN
```

```
  Query Match             4.3%;  Score 46;  DB 12;  Length 1007;
  Best Local Similarity   37.9%;  Pred. No. 2.1;
  Matches  153;  Conservative   45;  Mismatches  205;  Indels    1;  Gaps    1;

Qy    125 attccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagc 184
          |||| || |   ||| | | |: |    | | |||||   |    |   |||| ||
Db      4 ATTCTAAAGTTTATCGGAATAWAATACCACAAGAAAAGAAAGTCCTGGTTGATGCTGTTG 63

Qy    185 tgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaagg 244
          | ||  | ||    :| | :| :     :|| | : | ||: |||| || ::|||||
Db     64 TTAATAAATTAAAGMAAAAWGGWAAWAAWGGWAAWAAWGAWAAAAAAAAAAAAAWWAAAGG 123

Qy    245 tgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagt 304
           :    | | |  |      ||    || ||| ::     :: ||| | |   ||
Db    124 CAWTCGGTCTCTGGAGGAAAAAAAAAAATWAAAAWWTAAAWMMAAAAAATATAAAGTTTA 183

Qy    305 atatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagc 364
          | |  | || |   |   |     | || |:  |:   |  ||:||  |:       |
Db    184 AAAAAGWAAAGAGAAATGAWAAAWTAAAMAWMMAWAAACAMGTRCATGTWCMWGGTGATG 243

Qy    365 ctaatagcaatgaa-ataggaaaagtggagtctgcaagtccggaagacgtaagatatttt 423
           :: |:   || |: |:|  :|:|:|  :|  |  |  :    :| |  :|  | :  ||
Db    244 AWMGTRAWTATWAWGAWAAARAWARTAAWGGWTAAAGMWAWAAWAAAAAWAGAAGWGGTT 303

Qy    424 aaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaattcct 483
```

```
          | ||  |:  : | ||:: :| :||||    ||| ||: ||| :    |    ||:   :
Db    304 ATAAWAAWAAWAAWAAWRAWAAWAAAAAAAAAAAAGCAMAAAAAAWAAAAAAMAAAAAWAAAW 363

Qy    484 gatgtagctacattaaattctttattcaatcaaataaagaatca 527
           |   :   | |  |||:: |     ||  ||||||| :|| |
Db    364 AAAAAMAWAAAACGAAAWWAATGTAAAAAAAAAATAAAAWATAA 407


RESULT  10
CNS00LJT
LOCUS       CNS00LJT                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR30P10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL068307
VERSION     AL068307.1  GI:4958538
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR30P10"
                     /note="end : T7"
BASE COUNT      507 a    148 c    112 g    171 t    163 others
ORIGIN
```

```
  Query Match             4.3%;  Score 46;  DB 12;  Length 1101;
  Best Local Similarity   36.8%;  Pred. No. 2.1;
  Matches  156;  Conservative   43;  Mismatches  225;  Indels    0;  Gaps    0;

Qy      5 aaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatga 64
          || | ||  ||  ||   || ||||:   |  |   ||  | |||  | ||  :  :  |
Db     93 AATAAAAAAAAAAAAAAAAAAAAAAAWAAAAAAAAATAAAAAAAAAAAAAAAARRAMAWA 152

Qy     65 aaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccg 124
          ||||  :| |||||  ||  |            |   |:  |      ||:   |  :
Db    153 AAAAAAWTATTTTAAAAAAAAAAAGAAAAAAAAAAAATMAAATTAAAMAAWAWWTAARAAA 212

Qy    125 attccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagc 184
          |  ||:    ||  :  :|| |||   |||  |||||  ||    |||| |    ||
Db    213 TTAACAMAAAAAAAAAWAWMAAACAAAAWAAAAAAAAAAAAAAAAAAAAAAATAAAAAMAAAA 272

Qy    185 tgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaagg 244
            |||        ||| ||| :      :   ||  |  || | ||   :|  | |||
Db    273 AAAAAATACAAAMAAAAACTTWGAAAAAWAAAATCAAAAAAAAGAAAAWWAAAAAAAAAA 332
```

```
Qy    245 tgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagt 304
Db    333 AWAAAWAAAWWTWWWWTWWWMAAWAAWAWAAWAWCAAAWMAAAAAAAAAAATAGAAAAAA 392

Qy    305 atatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagc 364
Db    393 AATAMAAAWATHTTTCHGAWAMATYTAAAAAWMAATAMAAMAAAAARAATWAAAMTAAAA 452

Qy    365 ctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatatttta 424
Db    453 CWWAAAAWAAAMACWAAAAAAWAAAGAAAAAAGAAWCMAAAAAWAWTWAAAWGWWACATM 512

Qy    425 aaac 428
Db    513 ATAC 516
```

RESULT 11
CNS005TE/c
LOCUS CNS005TE 997 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL060767
VERSION AL060767.1 GI:4943573
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 997)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR12K22"
/note="end : TET3"
BASE COUNT 89 a 99 c 13 g 258 t 538 others
ORIGIN

Query Match 4.2%; Score 45.8; DB 12; Length 997;
Best Local Similarity 17.4%; Pred. No. 2.3;
Matches 59; Conservative 129; Mismatches 151; Indels 0; Gaps 0;

```
Qy    135 gaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagattt 194
Db    892 RAARDARARARARARRRARAKAARARGARRRRGRRRRRGRAGRRRARGAGGRAGRRGR 833

Qy    195 cggaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatttat 254
```

```
Db    832 GRGRRRARARRRAGAARARARGRRARAARARARRRRAGRRRRRARGARGRRRRGGRRRGR 773

Qy    255 gttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaat 314
Db    772 GGRRRGRRRRAARRARARARAGARAAARRRRARRAGAAARRARRRRRRRRRRGAGARARR 713

Qy    315 gcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaa 374
Db    712 ARRRARRAARAGAGAARRRRGRRGRGAGARRRGRGRGRRRGAGARRRRRGRRMRTRRARR 653

Qy    375 tgaaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcct 434
Db    652 RRRARRGAGAARRRARGARRRRGRRGRRRRGRARGRRRRAGRGRRRGRRARGARRGRRR 593

Qy    435 gacaaaagaagtaaaagggcaaaccaataaattggcgag 473
Db    592 AGRRRRRRARAGAGARRGRRRRGRARRAARGAARRRRAR 554
```

RESULT 12
CNS010PW/c
LOCUS CNS010PW 816 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN04A18 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL099278
VERSION AL099278.1 GI:5610889
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 816)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .816
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN04A18"
/note="end : T7"
BASE COUNT 46 a 16 c 30 g 371 t 353 others
ORIGIN

Query Match 4.2%; Score 45.6; DB 12; Length 816;
Best Local Similarity 29.1%; Pred. No. 2.5;
Matches 111; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

```
Qy    143 aaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagatttcggaaaga 202
Db    626 AAACNCNANNAANNAAAAAANAAAAAAAAAAAAAAAACAAAAAAAANNNANNNNNNANNN 567

Qy    203 ctgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcatttatgttaactg 262
Db    566 NNAAAAANAANANNAAANANNNAAAAAANNAAANAAAAAACNNNNTNNNTNNTATANNAA 507

Qy    263 cgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaatgcttagac 322
```

```
Db    506 TNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNNNANNNNNNNAAAAANANANAN 447

Qy    323 aggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagcaatgaaatag 382

Db    446 NTNNTNNNAANANNNNAAAANNANANAANANNAAAAAAAAAANNNAAAAAAAAAAAAAAN 387

Qy    383 gaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcctgacaaaag 442

Db    386 AAANNNANNANNNANNAAANNNNNAANNNNNNAANNNNNNANNNAANNANACGNTNNNNN 327

Qy    443 aagtaaaagggcaaaccaataaattggcgagtgtaattcctgatgtagctacattaaatt 502

Db    326 NNNNAANNNANAAANNNNNANNNNNNNNNNNANAAANACCACNNNNNAACCGANNNANA 267

Qy    503 ctttattcaatcaaataaagaa 524

Db    266 CTACANANAAAAAAAAAAAAA 245


RESULT 13
BH137450
LOCUS       BH137450                965 bp    DNA     linear   GSS 07-AUG-2001
DEFINITION  ENTOI31TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH137450
VERSION     BH137450.1  GI:15096511
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 770.
FEATURES             Location/Qualifiers
     source          1. .965
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      523 a     52 c    202 g    188 t
ORIGIN


  Query Match             4.2%;  Score 45.6;  DB 12;  Length 965;
  Best Local Similarity   46.1%;  Pred. No. 2.5;
  Matches  153;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy    134 ggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagctgaaagatt 193

Db    195 GAAAAGAAGAAATAAAAAGAATATTAGAAGAATATCATATTGATATTACAATCAAAGAAT 254

Qy    194 tcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaaggtgtcattta 253

Db    255 ATATAAAAACAGTTAAAAGAAGTAAAGGGATTAAAAAAGAAATCATAGAATACATTAATA 314

Qy    254 tgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagtatatcggaa 313

Db    315 AGAAAGTAGGAGAATTGATTAAAGAAGAAGATAATATAAAAAATGAAAAGATTAGAGAGA 374

Qy    314 tgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagcctaatagca 373

Db    375 TTATTAGAAAGAAAAAGAAGGAAGAGATTGATGTAGAATACATAGAAGAGTTAATAAAGA 434

Qy    374 atgaaataggaaaagtggagtctgcaagtccggaagacgtaagatattttaaaacgatcc 433

Db    435 AAAGTAAAAGAAAAGACATCAAAGAGTGTAGATACATTGGATATTATTTCAATAAAGAAA 494

Qy    434 tgacaaaagaagtaaaagggcaaaccaataaa 465

Db    495 TATGTGAAGAAATAATAAGATATATTGATGAA 526


RESULT 14
CNS01UPZ
LOCUS       CNS01UPZ                581 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            197M17 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL168128
VERSION     AL168128.1  GI:7806185
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 581)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 581)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="197M17"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG197AG09SP1~end :
                     PUC-Ori"
```

```
BASE COUNT      417 a      7 c     58 g     31 t     68 others
ORIGIN


  Query Match             4.2%;  Score 45;  DB 12;  Length 581;
  Best Local Similarity   38.4%;  Pred. No. 3.3;
  Matches  147;  Conservative   20;  Mismatches  216;  Indels    0;  Gaps    0;

Qy       5 aaaataaccaaccaacttaacaaaaactcaccattaaactacaaattacaattattatga 64
           |||| |   ||  ||   || |||||   |  |  ||| || ||| :| ||  |  |  |
Db     159 AAAAAATAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAAAAATAAAAAAWAAAAAAAAAAAA 218

Qy      65 aaaatctttttttatcaatgatggcctttgtgaccgtcttaacttttaattcctgtgccg 124
           ||||         |  ||  |           |       ||     |
Db     219 AAAAAAAAAAAAAAAAAAAAAAAAANNNNNNANANNNNAAAAAAAAAAAANNNNANNNANNN 278

Qy     125 attccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagc 184
           |    |   ||  |  |||  || | |||  |||||  || :  || | |   | |
Db     279 ANNNNANAAAAAAAAAAAAAAAAAAAAGWAAAAAAAAAAAAAAAAWAAAATTAAAAATGAAA 338

Qy     185 tgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaagg 244
           | |:| |:::   |||  |: ||   :|   ||: |   : ||||   || || |||
Db     339 TAAWAWAWWWAAAAAAATCWATAWWTAWATAAATWAAWACMAAAAAAAAAAAGAAAAAAA 398

Qy     245 tgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagt 304
           :: :|  |    |:    | |:    |  ||   |     |  ||| |||:  || | |
Db     399 ARWAWTAAAAAGAAWARAGGAARGAGRAAAAATAAATAAAAAARAAAAAAARRAAGGAAT 458

Qy     305 atatcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttcttaaagc 364
           |:|    ||     | | |       || |:: ||     :  | | |     | |||
Db     459 AWAAAAAAAAAAAAAAAAAWAGAAAAAAAAWWAAAAARAAWWAAAAAAGAAGATAAAAAA 518

Qy     365 ctaatagcaatgaaataggaaaa 387
           || |  |   ||| |  ||||
Db     519 TAAAAAAAATAAAAAAAAAAAA 541


RESULT  15
CNS07E8R/c
LOCUS       CNS07E8R                 964 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T7 end of clone XBD0AA002B08 of library XBD0AA from strain CBS 94
            of Candida tropicalis, genomic survey sequence.
ACCESSION   AL441457
VERSION     AL441457.1  GI:12224683
KEYWORDS    GSS.
SOURCE      Candida tropicalis.
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 964)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
REFERENCE   2  (bases 1 to 964)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
REFERENCE   3  (bases 1 to 964)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Candida tropicalis"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="XBD0AA002B08"
                     /clone_lib="XBD0AA"
                     /note="end : T7"
BASE COUNT      269 a    106 c     78 g    425 t     86 others
ORIGIN


  Query Match             4.2%;  Score 45;  DB 12;  Length 964;
  Best Local Similarity   36.8%;  Pred. No. 3.5;
  Matches  152;  Conservative   51;  Mismatches  206;  Indels    4;  Gaps    1;

Qy     125 attccaacgggaatcaggaaatcaacggaaaggaaaaactaagtgtaaatgattctaagc 184
           |||   |   |||  |  |||  || || ||  |||||  ||    |||  |   :|||
Db     585 ATTAATAAAAGAAAAAAAAAAAAAAAGGGAAAAAAAAAAAAAAAAAAAAAAAAAAWAAGG 526

Qy     185 tgaaagatttcggaaagactgtaccggtagggatagacgaagaaaacggaatgataaagg 244
             ||| |||    |:| |    |     |   | |:|  || |  | | |  | | ||| :
Db     525 GAAAAAATTGGWAAWAAAAAAGAAAAWRAAAAAAARAAAAAAATGAGGTGAAAAGAAAAR 466

Qy     245 tgtcatttatgttaactgcgcaattctatgaaattaagccgaccaaagaaaatgagcagt 304
           |   |   |    || :    ||:: | | :  : |:   ||  ||| ||||| :  | :
Db     465 TAAAAAAAAAAARAAAWAAAAAAAWWTTGTTRGTWAARAWWGAAAAAAWAAAATAWAAAAW 406

Qy     305 at----atcggaatgcttagacaggctgttaagaatgaatctcctgtacacattttctta 360
           ::    |    ||:    | | |     : || ||: :|: |    :  | |  :: | |
Db     405 WWRARAARAAAAAWATGAAWARAAAACCWAAAAAAWAWAWTTAAAAWTAAAARAWWTTAA 346

Qy     361 aagcctaatagcaatgaaataggaaaagtggagtctgcaagtccggaagacgtaagatat 420
           |     :|| |     ||::     :| || | | |  ||      :| :  |   |:|
Db     345 ATTTAAMATTGTCTRWAAWWTWTGTWAWTGTAWTGTAAAAAAAAAARAAWRWTTGWAWAA 286

Qy     421 tttaaaacgatcctgacaaaagaagtaaaagggcaaaccaataaattggcgagtgtaatt 480
           ::   |:  ::  : | |||:   |:|::|    |||  |: |:| |    |   :::|
Db     285 WWATTAWWAWWWAWAAAAAAWTGTGWAWWAAWARAAAAAAWAAWAATAAAAAAAAAWWWT 226

Qy     481 cctgatgtagctacattaaattctttattcaatcaaataaagaatcagtcttg 533
             :||||| :: :|:||:||    ||  ||  ||| ||| ||  |   ||:
Db     225 TARARTGTAGTWWWWTWAAWTTAGAAATAAAAAAAAAAAAAAAAAAARATTR 173


Search completed: June 27, 2002, 19:22:12
Job time: 10620 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:47:22 ; Search time 84.27 Seconds
(without alignments)
421.783 Million cell updates/sec

Title: US-09-727-769A-8
Perfect score: 1679
Sequence: 1 MKNLFLSMMAFVTVLTFNSC.....FSLLSGCSPSPAPDVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1679	100.0	320	22	AAB84387	Amino acid sequenc
2	1292.5	77.0	319	21	AAY44583	Chryseobacterium g
3	991	59.0	185	22	AAB84386	Amino acid sequenc
4	874	52.1	185	21	AAY44582	Chryseobacterium g
5	116.5	6.9	884	15	AAR63793	Bacillus cereus 10
6	112.5	6.7	852	17	AAR91246	VIP1A(a) protein w
7	112.5	6.7	852	18	AAW19516	Maize optimised-B.
8	112.5	6.7	852	19	AAW46727	Maize optimised VI
9	112.5	6.7	880	19	AAW60224	Bacillus thuringie
10	112.5	6.7	881	21	AAY59277	MIS toxin from B.
11	112.5	6.7	884	17	AAR91239	B. cereus VIP1A(a)
12	112.5	6.7	884	18	AAW19509	B. cereus VIP1A(a)
13	112.5	6.7	884	19	AAW46712	100 kDa VIP1A(a) t
14	112.5	6.7	1338	17	AAR91247	VIP2A(a)-VIP1A(a)
15	112.5	6.7	1338	18	AAW19520	Maize optimised-B.
16	112.5	6.7	1338	19	AAW46731	VIP2A(a)/VIP1A(a)
17	112.5	6.7	1346	17	AAR91245	VIP2A(a) and VIP1A
18	112.5	6.7	1346	18	AAW19513	B. cereus VIP1A(a)
19	112.5	6.7	1346	19	AAW46723	VIP1A(a)/VIP2A(a)
20	108	6.4	20	22	AAB84385	Internal peptide o
21	99	5.9	594	21	AAG52451	Arabidopsis thalia
22	99	5.9	594	21	AAG52479	Arabidopsis thalia
23	96	5.7	20	22	AAB84384	N-terminal peptide
24	95	5.7	1541	11	AAR07304	IgA1 protease. Ha
25	93	5.5	264	21	AAG25396	Arabidopsis thalia
26	93	5.5	292	21	AAG25395	Arabidopsis thalia
27	93	5.5	294	21	AAG25394	Arabidopsis thalia
28	93	5.5	1350	22	AAG63963	Amino acid sequenc
29	92.5	5.5	456	18	AAW36050	Hybrid Marek's dis
30	91.5	5.4	450	19	AAW98762	H. pylori GHPO 109
31	91.5	5.4	615	15	AAR63230	Mycoplasma gallise
32	91.5	5.4	1086	18	AAW36051	Hybrid Marek's dis
33	91	5.4	812	21	AAY73356	HTRM clone 2076520
34	91	5.4	848	21	AAY88565	Human NCAM 140kD i
35	91	5.4	1185	13	AAR22675	Collagen binding p
36	90	5.4	477	22	AAB30815	Amino acid sequenc
37	89.5	5.3	610	15	AAR63229	Mycoplasma gallise
38	89.5	5.3	1338	14	AAR41731	High molecular wei
39	89.5	5.3	1598	18	AAW30291	Non-typeable Haemo
40	89.5	5.3	2500	21	AAB18272	Plasmodium falcipa
41	89	5.3	661	22	ABG08463	Novel human diagno
42	88.5	5.3	512	19	AAW31554	Collagen binding p
43	88.5	5.3	1252	16	AAR80530	B. sphaericus SLP.
44	88	5.2	458	22	ABB61099	Drosophila melanog
45	88	5.2	761	22	ABB58195	Drosophila melanog

ALIGNMENTS

RESULT 1
AAB84387
ID AAB84387 standard; Protein; 320 AA.
XX
AC AAB84387;
XX
DT 22-AUG-2001 (first entry)
XX
DE Amino acid sequence of a protein-deamidating enzyme.
XX
KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW bakery; confectionery.
XX
OS Cryseobacterium sp.
XX
FH Key Location/Qualifiers
FT Region 1..135
FT /note= "prepro region"
XX
PN EP1106696-A1.
XX
PD 13-JUN-2001.
XX
PF 04-DEC-2000; 2000EP-0310768.
XX
PR 03-DEC-1999; 99JP-0345044.
XX
PA (AMAN-) AMANO ENZYME INC.
XX
PI Yamaguchi S;
XX
DR WPI; 2001-376907/40.
DR N-PSDB; AAF90281.
XX

```
PT   New enzyme for use in e.g. bakery has an ability to deamidate amido
PT   groups in a protein -
XX
PS   Example 11; Page 24; 43pp; English.
XX
CC   The present sequence represents a protein-deamidating enzyme from
CC   Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
CC   groups in a protein by directly acting upon the amido groups without
CC   cutting peptide bonds and without cross-linking the protein. The enzyme
CC   thus reduces the mineral sensitivity of the protein and increases the
CC   soluble mineral content in the protein-mineral solution, improving the
CC   absorption of minerals in the human body. The enzyme reduces or removes
CC   the toxicity and allergenic property of the protein in a food
CC   e.g. allergy. The enzyme is useful for the improvement of dough in the
CC   field of bakery and confectionery e.g. for the production of crackers,
CC   biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC   protein products, in various food articles e.g. meat or fish products
CC   and noodles; and for improving functionality of plant or animal protein.
XX
SQ   Sequence   320 AA;


  Query Match             100.0%;  Score 1679;  DB 22;  Length 320;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-150;
  Matches  320;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mknlflsmmafvtvltfnscadsngnqeingkeklsvndsklkdfgktvpvgideengmi 60

Qy      61 KVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 kvsfmltaqfyeikptkeneqyigmlrqavknespvhiflkpnsneigkvesaspedvry 120

Qy     121 FKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYAR 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 fktiltkevkgqtnklasvipdvatlnslfnqiknqscgtstasspcitfrypvdgcyar 180

Qy     181 AHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPS 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ahkmrqilmnngydcekqfvygnlkastgtccvawsyhvailvsyknasgvtekriidps 240

Qy     241 LFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 lfssgpvtdtawrnacvntscgsasvssyantagnvyyrspsnsylydnnlintncvltk 300

Qy     301 FSLLSGCSPSPAPDVSSCGF 320
           ||||||||||||||||||||
Db     301 fsllsgcspspapdvsscgf 320


RESULT   2
AAY44583
ID   AAY44583 standard; Protein; 319 AA.
XX
AC   AAY44583;
XX
DT   04-APR-2000  (first entry)
XX
DE   Chryseobacterium gleum protein-deamidating prepro-enzyme.
XX
KW   Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;
KW   amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW   protein engineering; surface hydrophobicity; toxicity; allergic;
KW   mineral sensitivity; calcium; absorption; mineral enhancing agent.
XX
OS   Chryseobacterium gleum 'JCM 2410'.
XX
FH   Key             Location/Qualifiers
FT   Region          1..134
FT                   /label= Prepro_region
FT                   /note= "Comprises the Pre and Pro region"
FT   Peptide         1..21
FT                   /label= Signal_peptide
FT                   /note= "Corresponds to the Pre region"
FT   Peptide         22..134
FT                   /note= "Corresponds to the Pro region"
FT   Protein         135..319
FT                   /label= Mature_protein_deamidating_enzyme
FT                   /note= "Deamidates amido groups in a protein"
XX
PN   EP976829-A2.
XX
PD   02-FEB-2000.
XX
PF   04-JUN-1999;   99EP-0304367.
XX
PR   04-JUN-1998;   98JP-0173940.
XX
PA   (AMAN ) AMANO PHARM KK.
XX
PI   Yamaguchi S,  Matsuura A;
XX
DR   WPI; 2000-118552/11.
DR   N-PSDB; AAZ49495.
XX
PT   New enzyme for modifying and improving the function of proteins and/or
PT   peptides has deamidating activity without causing cross linking -
XX
PS   Example 26; Page 26; 57pp; English.
XX
CC   The present amino acid sequence is the protein-deamidating
CC   prepro-enzyme, isolated from a new strain of soil bacterium,
CC   Chryseobacterium gleum JCM 2410. This enzyme exerts the deamidation
CC   activity by directly acting upon side chain amido groups in the protein
CC   in bonded state and releasing side chain carboxyl groups and ammonia. It
CC   can deaminate high molecular weight proteins, without cross linking and
CC   cleavage of peptide bonds, to improve protein function. This sequence is
CC   used for protein engineering, to cause an increase in surface
CC   hydrophobicity and improve the function of a plant or animal protein. It
CC   can also be used to remove or reduce toxicity or allergic property of
CC   proteins in food, decrease mineral sensitivity of protein, to allow
CC   greater absorption into the body and to solubilise calcium for use in
CC   drinks and mineral enhancing agents.
XX
SQ   Sequence   319 AA;


  Query Match             77.0%;  Score 1292.5;  DB 21;  Length 319;
  Best Local Similarity   74.7%;  Pred. No. 1.7e-113;
  Matches  239;  Conservative   39;  Mismatches   41;  Indels    1;  Gaps    1;

Qy       1 MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMI 60
           ||   |||||||:|:||:|:||: ||: |   | | |:  :|||||||||||::|:|
Db       1 mkkfllsmmafvtilsfnacsdssanqdpnlvakes-nevamkdfgktvpvgiekedgkf 59

Qy      61 KVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRY 120
           |:|||:||| |||  :|||  || |:||||:||:|||:||| |:|:| ||| |: :|:||
Db      60 kisfmvtaqpyeiadskenagyismirqavenetpvhvflkvntnkiakvekatdddiry 119

Qy     121 FKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYAR 180
           ||::  |: :|::||  |||||:||||||| |||||:|||||||||||||||||||||||
Db     120 fksvfnkqergesnkavsvipdlatlnslftqiknqacgtstasspcitfrypvdgcyar 179

Qy     181 AHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPS 240
           ||||||||:| |||||||||||||:||||||||:| ||||||||:|||||: ||||||||
Db     180 ahkmrqillnagydcekqfvygnlrastgtccvswvyhvailvsfknasgivekriidps 239

Qy     241 LFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTK 300
           |||||||||:||| || |||||||||||||||||||||| | ||||| :||||||
Db     240 lfssgpvtdsawraactntscgsasvssyantagnvyyrspsgsllydnnyvntncvlni 299

Qy     301 FSLLSGCSPSPAPDVSSCGF 320
           || |||||||||| |:||||
Db     300 fsslsgcspspapsvascgf 319
```

```
RESULT   3
AAB84386
ID   AAB84386 standard; Protein; 185 AA.
XX
AC   AAB84386;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Amino acid sequence of a protein-deamidating enzyme.
XX
KW   Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW   bakery; confectionery.
XX
OS   Cryseobacterium sp.
XX
PN   EP1106696-A1.
XX
PD   13-JUN-2001.
XX
PF   04-DEC-2000; 2000EP-0310768.
XX
PR   03-DEC-1999;   99JP-0345044.
XX
PA   (AMAN-) AMANO ENZYME INC.
XX
PI   Yamaguchi S;
XX
DR   WPI; 2001-376907/40.
DR   N-PSDB; AAF90280.
XX
PT   New enzyme for use in e.g. bakery has an ability to deamidate amido
PT   groups in a protein  -
XX
PS   Claim 4; Page 22; 43pp; English.
XX
CC   The present sequence represents a protein-deamidating enzyme from
CC   Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
CC   groups in a protein by directly acting upon the amido groups without
CC   cutting peptide bonds and without cross-linking the protein. The enzyme
CC   thus reduces the mineral sensitivity of the protein and increases the
CC   soluble mineral content in the protein-mineral solution, improving the
CC   absorption of minerals in the human body. The enzyme reduces or removes
CC   the toxicity and allergenic property of the protein in a food
CC   e.g. allergy. The enzyme is useful for the improvement of dough in the
CC   field of bakery and confectionery e.g. for the production of crackers,
CC   biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC   protein products, in various food articles e.g. meat or fish products
CC   and noodles; and for improving functionality of plant or animal protein.
XX
SQ   Sequence   185 AA;

  Query Match             59.0%;  Score 991;  DB 22;  Length 185;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-85;
  Matches  185;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    136 LASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDC 195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 lasvipdvatlnslfnqiknqscgtstasspcitfrypvdgcyarahkmrqilmnngydc 60

Qy    196 EKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNA 255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ekqfvygnlkastgtccvawsyhvailvsyknasgvtekriidpslfssgpvtdtawrna 120

Qy    256 CVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDV 315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cvntscgsasvssyantagnvyyrspsnsylydnnlintncvltkfsllsgcspspapdv 180

Qy    316 SSCGF 320
          |||||
Db    181 sscgf 185
```

```
RESULT   4
AAY44582
ID   AAY44582 standard; Protein; 185 AA.
XX
AC   AAY44582;
XX
DT   04-APR-2000  (first entry)
XX
DE   Chryseobacterium gleum protein-deamidating enzyme.
XX
KW   Protein deamidating enzyme; soil bacterium; deamidation activity;
KW   amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW   protein engineering; surface hydrophobicity; toxicity; allergic;
KW   mineral sensitivity; calcium; absorption; mineral enhancing agent.
XX
OS   Chryseobacterium gleum 'JCM 2410'.
XX
PN   EP976829-A2.
XX
PD   02-FEB-2000.
XX
PF   04-JUN-1999;   99EP-0304367.
XX
PR   04-JUN-1998;   98JP-0173940.
XX
PA   (AMAN ) AMANO PHARM KK.
XX
PI   Yamaguchi S,  Matsuura A;
XX
DR   WPI; 2000-118552/11.
DR   N-PSDB; AAZ49494.
XX
PT   New enzyme for modifying and improving the function of proteins and/or
PT   peptides has deamidating activity without causing cross linking -
XX
PS   Claim 5; Page 24; 57pp; English.
XX
CC   The present amino acid sequence is the protein-deamidating enzyme,
CC   isolated from a new strain of soil bacterium, Chryseobacterium gleum
CC   JCM 2410. This enzyme exerts the deamidation activity by directly acting
CC   upon side chain amido groups in the protein in bonded state and
CC   releasing side chain carboxyl groups and ammonia. It can deaminate high
CC   molecular weight proteins, without cross linking and cleavage of peptide
CC   bonds, to improve protein function. This sequence is used for protein
CC   engineering, to cause an increase in surface hydrophobicity and improve
CC   the function of a plant or animal protein. It can also be used to remove
CC   or reduce toxicity or allergic property of proteins in food, decrease
CC   mineral sensitivity of protein, to allow greater absorption into the body
CC   and to solubilise calcium for use in drinks and mineral enhancing agents.
XX
SQ   Sequence   185 AA;

  Query Match             52.1%;  Score 874;  DB 21;  Length 185;
  Best Local Similarity   87.4%;  Pred. No. 2.4e-74;
  Matches  160;  Conservative   10;  Mismatches   13;  Indels    0;  Gaps    0;

Qy    138 SVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEK 197
          |||||:|||||| |||||:|||||||||||||||||||||||||||||:| ||||||
Db      3 svipdlatlnslftqiknqacgtstasspcitfrypvdgcyarahkmrqillnagydcek 62

Qy    198 QFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACV 257
          |||||||:|||||||:| |||||||:|||||: ||||||||||||||||:||| ||
Db     63 qfvygnlrastgtccvswvyhvailvsfknasgivekriidpslfssgpvtdsawraact 122

Qy    258 NTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSS 317
          |||||||||||||||||||||| | ||||| :|||||  || ||||||||| |:|
Db    123 ntscgsasvssyantagnvyyrspsgsllydnnyvntncvlnifsslsgcspspapsvas 182

Qy    318 CGF 320
          |||
Db    183 cgf 185
```

```
RESULT   5
AAR63793
ID   AAR63793 standard; Protein; 884 AA.
XX
AC   AAR63793;
XX
DT   12-JUL-1995  (first entry)
XX
DE   Bacillus cereus 100 kDa vegatative insecticidal protein-1.
XX
KW   Vegetative insecticidal protein-1; VIP-1; Bacillus cereus;
KW   pesticide; Diabrotica virgifera virgifera; insecticide.
XX
OS   Bacillus cereus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 217
FT                   /label= Lys, Gln
FT   Misc-difference 245
FT                   /label= Gln, His
XX
PN   WO9421795-A.
XX
PD   29-SEP-1994.
XX
PF   23-MAR-1994;   94WO-US03131.
XX
PR   25-MAR-1993;   93US-0037057.
XX
PA   (CIBA ) CIBA GEIGY AG.
XX
PI   Carr B,  Desai N,  Kostichka N,  Koziel MG,  Mullins MA;
PI   Nye GJ,  Warren GW;
XX
DR   WPI; 1994-317015/39.
DR   N-PSDB; AAQ74679.
XX
PT   Novel pesticidal proteins and Bacillus strains - e.g. useful for
PT   control of Diabrotica virgifera virgifera
XX
PS   Claim 18; Page 63; 108pp; English.
XX
CC   AAQ74679 encodes AAR63793 the 100 kDa vegetative insecticidal protein-1
CC   (VIP-1), also claimed is the protein AAR63795 which enhances the
CC   pesticidal activity of VIP-1. The protein peptide combination can be
CC   used in a variety of systems for controlling plant and non-plant
CC   pests, including insects, fungi, bacteria, nematodes, protozoan
CC   pathogens and animal-parasitic liver flukes. However it is esp.
CC   useful in the control of Diabrotica virgifera virgifera.
XX
SQ   Sequence   884 AA;


  Query Match             6.9%;  Score 116.5;  DB 15;  Length 884;
  Best Local Similarity   23.5%;  Pred. No. 0.088;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     114 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 166

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     167 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqxmkr 220

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :       | | |   |:|  :  : |   |
Db     221 eidedtdtdgdsipdlweengytixnriavkwddslaskgytkfvsnples--htvgdpy 278

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               |  | : ::|  :               ||  :  |   |          | :

Db     279 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 333

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
             || |  | : :  | |   | |             |  :  |||  : : : | |
Db     334 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 393

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     394 rynnvgtgaiydvkpttsfvlnndtiat 421


RESULT   6
AAR91246
ID   AAR91246 standard; Protein; 852 AA.
XX
AC   AAR91246;
XX
DT   15-AUG-1996  (first entry)
XX
DE   VIP1A(a) protein with signal peptide removed.
XX
KW   Pesticide; insecticide; biological control agent; Lepidoptera;
KW   Coleoptera; transgenic plant; maize; Zea mays; insect resistance;
KW   VIP1; Bacillus cereus; western corn rootworm;
KW   Diabrotica vergifera vergifera.
XX
OS   Synthetic.
XX
PN   WO9610083-A1.
XX
PD   04-APR-1996.
XX
PF   27-SEP-1995;   95WO-EP03826.
XX
PR   05-JUN-1995;   95US-0463483.
PR   28-SEP-1994;   94US-0314594.
XX
PA   (CIBA ) CIBA GEIGY AG.
XX
PI   Carr B,  Desai NM,  Duck NB,  Estruch JJ,  Kostichka K;
PI   Koziel MG,  Mullinsma,  Nye GJ,  Warren GW;
XX
DR   WPI; 1996-200921/20.
DR   N-PSDB; AAT13945.
XX
PT   Bacillus strain producing insecticidal protein during vegetative
PT   growth - used in the control of Lepidoptera and Coleoptera pests
XX
PS   Claim 31; Page 183-187; 242pp; English.
XX
CC   A version (AAR91246) of the Bacillus cereus strain AB78 VIP1A(a)
CC   insect-specific protein (see also AAR91239) lacks the Bacillus signal
CC   peptide and is the product of a synthetic gene (AAT13945) having codon
CC   usage optimised for expression in maize.  VIP1A(a) shows activity
CC   against western corn rootworm.  The synthetic gene can be fused to a
CC   synthetic gene (AAT13946) coding for the VIP2A(a) auxiliary protein.
CC   The resulting construct (AAT13947) codes for a VIP2A(a)-VIP1A(a) fusion
CC   protein (AAR91247) that can be expressed in maize.
XX
SQ   Sequence   852 AA;


  Query Match             6.7%;  Score 112.5;  DB 17;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.2;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db      82 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     135 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqqmkr 188
```

```
Qy    128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
          |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db    189 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 246

Qy    179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
            |  | : ::|  :               ||  :  |   |         | :
Db    247 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 301

Qy    223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |             |  :  |||  : : : | |
Db    302 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 361

Qy    273 ------AGNVYYRSPSNSYLYDNNLINT 294
                | :|   |: |:: :|: | |
Db    362 rynnvgtgaiydvkpttsfvlnndtiat 389
```

RESULT 7
AAW19516
ID AAW19516 standard; Protein; 852 AA.
XX
AC AAW19516;
XX
DT 15-APR-1998 (first entry)
XX
DE Maize optimised-B. cereus VIP1A(a) protein lacking secretion signal.
XX
KW Vegetative insecticidal protein; Bacillus cereus strain AB78; plant;
KW insect; Sesamia nonagrioides; maize; corn borer; toxin.
XX
OS Bacillus cereus.
OS Synthetic.
XX
PN WO9726339-A1.
XX
PD 24-JUL-1997.
XX
PF 23-DEC-1996; 96WO-EP05828.
XX
PR 15-JAN-1996; 96GB-0000786.
XX
PA (NOVS) NOVARTIS AG.
XX
PI Gay PB;
XX
DR WPI; 1997-385342/35.
DR N-PSDB; AAT74006.
XX
PT Protecting plants against insects of the genus Sesamia using
PT Bacillus toxic proteins - applied directly or expressed as
PT heterologous protein by the plant, also transgenic plants expressing
PT both Cry and VIP type toxins
XX
PS Claim 8; Page 102-106; 168pp; English.
XX
CC This is the amino acid sequence of the Bacillus cereus vegetative
CC insecticidal protein (VIP) 1A(a) which has been optimised for expression
CC in maize plants. The coding sequence was constructed such that the
CC Bacillus secretion signal sequence was deleted. The VIP protein can be
CC used in a new method for protecting plants, and their progeny, against
CC insects of the genus Sesamia by direct or indirect application to the
CC plant (or seed or growing area). The protein is especially useful to
CC protect maize plants against the Mediterranean corn borer
CC (S. nonagrioides).
XX
SQ Sequence 852 AA;

Query Match 6.7%; Score 112.5; DB 18; Length 852;
Best Local Similarity 23.5%; Pred. No. 0.2;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

```
Qy     16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
          |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     82 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 134

Qy     69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
          : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db    135 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqqmkr 188

Qy    128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
          |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db    189 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 246

Qy    179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
            |  | : ::|  :               ||  :  |   |         | :
Db    247 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 301

Qy    223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |             |  :  |||  : : : | |
Db    302 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 361

Qy    273 ------AGNVYYRSPSNSYLYDNNLINT 294
                | :|   |: |:: :|: | |
Db    362 rynnvgtgaiydvkpttsfvlnndtiat 389
```

RESULT 8
AAW46727
ID AAW46727 standard; Protein; 852 AA.
XX
AC AAW46727;
XX
DT 02-JUN-1998 (first entry)
XX
DE Maize optimised VIP1A(a) without the secretion signal.
XX
KW Vegetative insecticidal protein; VIP; expression; maize; protection;
KW plant; Ostrinia furnacalis; Asian Corn Borer; Cry toxin; VIP toxin;
KW recombinant; Bacillus thuringiensis; transgenic plant; resistance;
KW insect attack; Sesamia; maize; cereal crop.
XX
OS Synthetic.
OS Bacillus sp.
XX
PN WO9746105-A1.
XX
PD 11-DEC-1997.
XX
PF 27-MAY-1997; 97WO-EP02737.
XX
PR 06-JUN-1996; 96GB-0011777.
XX
PA (NOVS) NOVARTIS AG.
XX
PI Hunter B, Suwantaradon K, Uitdewilligen WPM;
XX
DR WPI; 1998-041787/04.
DR N-PSDB; AAV16181.
XX
PT Administration of Bacillus sp. toxin protein, especially Cry or
PT vegetative insecticidal protein (VIP) protein to plants - useful for
PT protection against attack by Asian Corn Borer (Ostrinia furnacalis)
XX
PS Claim 8; Pages 109-112; 175pp; English.
XX
CC The present sequence represents a vegetative insecticidal protein 1A(a)
CC (VIP1A(a)), with the Bacillus secretion signal removed. The sequence
CC is a synthetic sequence, in that the original codon usage of B. cereus
CC has been optimised for expression in maize. VIP toxins and genes are
CC used, especially inside recombinant B. cereus or B. thuringiensis
CC strains, to produce plants protected against Asian Borer pests.
CC Transgenic plants protected against Asian Corn Borer can be used to
CC produce seed and progeny also resistant to insect attack. Plants
CC expressing both a Cry-type and a VIP toxin gene can also protect

```
CC   against Sesamia pests. The method and compositions are especially used
CC   for protecting maize but may also be used to protect other cereal crops
CC   against Asian Corn Borer attack.
XX
SQ   Sequence   852 AA;


  Query Match             6.7%;  Score 112.5;  DB 19;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.2;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db      82 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |     | |  :  |   : :
Db     135 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqqmkr 188

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     189 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 246

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :                || :  |   |          | :
Db     247 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 301

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
            || |  | : :  | |   | |          |  :  |||  : : : | |
Db     302 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 361

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     362 rynnvgtgaiydvkpttsfvlnndtiat 389


RESULT   9
AAW60224
ID   AAW60224 standard; Protein; 880 AA.
XX
AC   AAW60224;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Bacillus thuringiensis insecticidal toxin 177C8.
XX
KW   Insecticide; pesticide; toxin; delta-endotoxin;
KW   biological control; lepidopteran; coleopteran.
XX
OS   Bacillus thuringiensis strain PS177C8 (NRRL B-21867).
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 253
FT                   /note= "encoded by YTA"
FT   Misc-difference 675
FT                   /note= "encoded by AC"
FT   Misc-difference 846
FT                   /note= "encoded by RAA"
XX
PN   WO9818932-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;   97WO-US19804.
XX
PR   30-OCT-1996;   96US-0029848.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Dullum CJ,  Feitelson JS,  Loewer D,  Muller-Cohn J;
PI   Narva KE,  Schmeits JL, Schnepf HE,  Schwab G,  Stamp L;
PI   Stockhoff BA;
XX
DR   WPI; 1998-272226/24.
DR   N-PSDB; AAV30307.
XX
PT   Bacillus thuringiensis isolates - used for producing pesticidal
PT   toxins and nucleotide sequences for control of lepidopterans and
PT   coleopterans
XX
PS   Claim 5; Page 81-84; 139pp; English.
XX
CC   This polypeptide comprises a novel soluble toxin of Bacillus
CC   thuringiensis (B.t.) strain PS177C8 (NRRL B-21867).  The toxin
CC   belongs to a novel family of B.t. toxins that have toxicity
CC   against non-mammalian pests.  Its amino acid sequence was deduced
CC   from a novel DNA fragment (see AAV30307) obtained by PCR from
CC   cellular genomic DNA of PS177C8.  Disclosed and claimed are novel
CC   B.t. isolates and toxins (see AAW60218-32) that have activity against
CC   lepidopteran and/or coleopteran pests, isolated genes, probes
CC   and primers (see AAV30288-321 and AAT99734-87) useful for production
CC   of the toxins and for the identification and characterisation of
CC   these toxins, and transformed hosts, particularly plant and
CC   bacterial hosts.  The invention provides 8 entirely new families of
CC   toxins from B.t. isolates.  The toxins have the additional ability
CC   to form pores in cell membranes, and can be used to facilitate
CC   entry of a second agent into a target cell.
XX
SQ   Sequence   880 AA;


  Query Match             6.7%;  Score 112.5;  DB 19;  Length 880;
  Best Local Similarity   23.5%;  Pred. No. 0.21;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     111 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 163

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |     | |  :  |   : :
Db     164 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqkmkr 217

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     218 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 275

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :                || :  |   |          | :
Db     276 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 330

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
            || |  | : :  | |   | |          |  :  |||  : : : | |
Db     331 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 390

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     391 rynnvgtgaiydvkpttsfvlnndtiat 418


RESULT  10
AAY59277
ID   AAY59277 standard; Protein; 881 AA.
XX
AC   AAY59277;
XX
DT   18-APR-2000  (first entry)
XX
DE   MIS toxin from B. thuringiensis strain PS177C8.
XX
KW   Bacillus thuringiensis; toxin; endotoxin; pesticide; plant pest;
KW   lepidopterans; cleopterans.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
```

```
FT   Misc-difference 846
FT                   /note= "unknown"
XX
PN   WO9957282-A2.
XX
PD   11-NOV-1999.
XX
PF   06-MAY-1999;   99WO-US09997.
XX
PR   06-MAY-1998;   98US-0073898.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Feitelson JS,  Schnepf HE,  Narva KE,  Stockhoff BA,  Schmeits J;
PI   Loewer D,  Dullum CJ,  Muller-Cohn J,  Stamp L,  Morrill G;
PI   Finstad-Lee S;
XX
DR   WPI; 2000-096811/08.
DR   N-PSDB; AAZ58786.
XX
PT   New polynucleotides encoding pesticidally active proteins, useful for
PT   transforming plants for controlling pests
XX
PS   Disclosure; Page 68-71; 104pp; English.
XX
CC   The invention relates to novel B. thuringiensis isolates, and genes
CC   encoding pesticidal toxins which are toxic to non-mammalian pests. The
CC   genes are useful in the control of non-mammalian pests and especially
CC   plant pests (e.g. lepidopterans and/or cleopterans). The polynucleotides
CC   are useful for transforming plants for controlling plant pests; for
CC   designing primers and probes useful for the identification and
CC   characterization of genes which encode pesticidal toxins. The present
CC   sequence represents a B.t. toxin.
XX
SQ   Sequence   881 AA;


  Query Match          6.7%;  Score 112.5;  DB 21;  Length 881;
  Best Local Similarity   23.5%;  Pred. No. 0.21;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     111 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 163

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     164 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqkmkr 217

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     218 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 275

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
           |  | : ::|  :                 ||  :  |   |          | :
Db     276 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 330

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
            || |  | : :  | |   | |              |  :  |||  : : : | |
Db     331 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 390

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     391 rynnvgtgaiydvkpttsfvlnndtiat 418


RESULT  11
AAR91239
ID   AAR91239 standard; Protein; 884 AA.
XX
AC   AAR91239;
XX
DT   14-AUG-1996  (first entry)
XX
DE   B. cereus VIP1A(a) insect-specific protein.
XX
KW   Pesticide; insecticide; biological control agent; Lepidoptera;
KW   Coleoptera; transgenic plant; maize; insect resistance;
KW   western corn rootworm; Diabrotica virgifera virgifera; VIP.
XX
OS   Bacillus cereus strain AB78 (NRRL B-21058).
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..33
FT                    /label= Sig_peptide
XX
PN   WO9610083-A1.
XX
PD   04-APR-1996.
XX
PF   27-SEP-1995;   95WO-EP03826.
XX
PR   05-JUN-1995;   95US-0463483.
PR   28-SEP-1994;   94US-0314594.
XX
PA   (CIBA ) CIBA GEIGY AG.
XX
PI   Carr B,  Desai NM,  Duck NB,  Estruch JJ,  Kostichka K;
PI   Koziel MG,  Mullinsma,  Nye GJ,  Warren GW;
XX
DR   WPI; 1996-200921/20.
DR   N-PSDB; AAT13940.
XX
PT   Bacillus strain producing insecticidal protein during vegetative
PT   growth - used in the control of Lepidoptera and Coleoptera pests
XX
PS   Claim 15; Page 121-124; 242pp; English.
XX
CC   Insect-specific protein VIP1A(a) (AAR91239) of Bacillus cereus AB78
CC   shows activity against Diabrotica spp. pests such as the
CC   western corn rootworm.  It is encoded by the VIP1A(a) gene (AAT13940)
CC   isolated from a cosmid clone of AB78.  VIP1A(a) can be expressed in
CC   e.g. bacterial hosts to provide biological control agents having
CC   increased activity or target range, or can be expressed in transgenic
CC   plants, esp. maize, to improve insect resistance.  It is preferably
CC   expressed as a fusion protein (see also AAR91245) with auxiliary protein
CC   VIP2A(a) (AAR91238).
XX
SQ   Sequence   884 AA;


  Query Match          6.7%;  Score 112.5;  DB 17;  Length 884;
  Best Local Similarity   23.5%;  Pred. No. 0.21;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     114 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 166

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     167 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqkmkr 220

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     221 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 278

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
           |  | : ::|  :                 ||  :  |   |          | :
Db     279 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 333

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
            || |  | : :  | |   | |              |  :  |||  : : : | |
Db     334 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 393

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
```

```
                | :|   |: |:: :|: | |
Db    394 rynnvgtgaiydvkpttsfvlnndtiat 421


RESULT  12
AAW19509
ID   AAW19509 standard; Protein; 884 AA.
XX
AC   AAW19509;
XX
DT   15-APR-1998  (first entry)
XX
DE   B. cereus VIP1A(a) protein sequence.
XX
KW   Vegetative insecticidal protein; Bacillus cereus strain AB78; plant;
KW   insect; Sesamia nonagrioides; maize; corn borer; toxin.
XX
OS   Bacillus cereus.
XX
PN   WO9726339-A1.
XX
PD   24-JUL-1997.
XX
PF   23-DEC-1996;   96WO-EP05828.
XX
PR   15-JAN-1996;   96GB-0000786.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Gay PB;
XX
DR   WPI; 1997-385342/35.
DR   N-PSDB; AAT73994.
XX
PT   Protecting plants against insects of the genus Sesamia using
PT   Bacillus toxic proteins - applied directly or expressed as
PT   heterologous protein by the plant, also transgenic plants expressing
PT   both Cry and VIP type toxins
XX
PS   Claim 8; Page 35-39; 168pp; English.
XX
CC   This is the amino acid sequence of the 100 kD vegetative insecticidal
CC   protein (VIP) 1A(a) from Bacillus cereus strain AB78.  The protein can
CC   be used in a new method for protecting plants, and their progeny, against
CC   insects of the genus Sesamia by direct or indirect application to the
CC   plant (or seed or growing area).  The protein is especially useful to
CC   protect maize plants against the Mediterranean corn borer
CC   (S. nonagrioides).
XX
SQ   Sequence   884 AA;


  Query Match             6.7%;  Score 112.5;  DB 18;  Length 884;
  Best Local Similarity   23.5%;  Pred. No. 0.21;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     114 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 166

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     167 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqkmkr 220

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     221 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 278

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
             |  | : ::|  :                ||  :  |   |           | :
Db     279 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 333

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
```

```
           || |  | : :  | |   | |             |  :  |||  : : : | |
Db    334 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 393

Qy    273 ------AGNVYYRSPSNSYLYDNNLINT 294
                | :|   |: |:: :|: | |
Db    394 rynnvgtgaiydvkpttsfvlnndtiat 421


RESULT  13
AAW46712
ID   AAW46712 standard; Protein; 884 AA.
XX
AC   AAW46712;
XX
DT   02-JUN-1998  (first entry)
XX
DE   100 kDa VIP1A(a) toxin of Bacillus cereus.
XX
KW   Vegetative insecticidal protein; VIP; expression; maize; protection;
KW   plant; Ostrinia furnacalis; Asian Corn Borer; Cry toxin; VIP toxin;
KW   recombinant; Bacillus thuringiensis; transgenic plant; resistance;
KW   insect attack; Sesamia; maize; cereal crop.
XX
OS   Bacillus cereus.
XX
PN   WO9746105-A1.
XX
PD   11-DEC-1997.
XX
PF   27-MAY-1997;   97WO-EP02737.
XX
PR   06-JUN-1996;   96GB-0011777.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Hunter B,  Suwantaradon K,  Uitdewilligen WPM;
XX
DR   WPI; 1998-041787/04.
DR   N-PSDB; AAV16166.
XX
PT   Administration of Bacillus sp. toxin protein, especially Cry or
PT   vegetative insecticidal protein (VIP) protein to plants - useful for
PT   protection against attack by Asian Corn Borer (Ostrinia furnacalis)
XX
PS   Claim 8; Pages 43-46; 175pp; English.
XX
CC   The present sequence represents a 100 kDa vegetative insecticidal
CC   protein 1A(a) (VIP1A(a)), and is derived from Bacillus cereus strain
CC   AB78. The protein is used in a method for protecting plants and their
CC   progeny against damage caused by Ostrinia furnacalis (Asian Corn Borer).
CC   The protein is directly or indirectly applied to the plant, plant seed
CC   or growing area of the plant. Cry toxins can also be used in the same
CC   way, in place of VIP toxins. The Cry or VIP toxins and genes are used,
CC   especially inside recombinant B. cereus or B. thuringiensis strains,
CC   to produce plants protected against Asian Borer pests. Transgenic plants
CC   protected against Asian Corn Borer can be used to produce seed and
CC   progeny also resistant to insect attack. Plants expressing both a
CC   Cry-type and a VIP toxin gene can also protect against Sesamia pests.
CC   The method and compositions are especially used for protecting maize but
CC   may also be used to protect other cereal crops against Asian Corn Borer
CC   attack.
XX
SQ   Sequence   884 AA;


  Query Match             6.7%;  Score 112.5;  DB 19;  Length 884;
  Best Local Similarity   23.5%;  Pred. No. 0.21;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     114 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 166
```

```
Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     167 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqkmkr 220

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     221 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 278

Qy     179 ARAHK-MRQILMNNGYD---------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :                ||  :  |   |           | :
Db     279 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 333

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
             || |  | : :  | |   | |              |  :  |||  : : : | |
Db     334 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 393

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     394 rynnvgtgaiydvkpttsfvlnndtiat 421
```

RESULT 14
AAR91247
ID AAR91247 standard; Protein; 1338 AA.
XX
AC AAR91247;
XX
DT 15-AUG-1996 (first entry)
XX
DE VIP2A(a)-VIP1A(a) protein fusion.
XX
KW Pesticide; insecticide; biological control agent; Lepidoptera;
KW Coleoptera; transgenic plant; maize; Zea mays; insect resistance;
KW VIP1; VIP2A; Bacillus cereus; western corn rootworm;
KW Diabrotica vergifera vergifera.
XX
OS Synthetic.
XX
PN WO9610083-A1.
XX
PD 04-APR-1996.
XX
PF 27-SEP-1995; 95WO-EP03826.
XX
PR 05-JUN-1995; 95US-0463483.
PR 28-SEP-1994; 94US-0314594.
XX
PA (CIBA) CIBA GEIGY AG.
XX
PI Carr B, Desai NM, Duck NB, Estruch JJ, Kostichka K;
PI Koziel MG, Mullinsma, Nye GJ, Warren GW;
XX
DR WPI; 1996-200921/20.
DR N-PSDB; AAT13947.
XX
PT Bacillus strain producing insecticidal protein during vegetative
PT growth - used in the control of Lepidoptera and Coleoptera pests
XX
PS Claim 31; Page 212-216; 242pp; English.
XX
CC A synthetic DNA (AAT13947), optimized for expression in maize,
CC encodes a fusion (AAR91247) between the VIP2A(a) (AAR91238) and
CC VIP1A(a) (AAR91239) proteins of Bacillus cereus strain AB78. It
CC is obtd. by fusing a synthetic VIP1A(a) gene (AAT13945) that lacks
CC a signal sequence, and a synthetic VIP2A(a) gene (AAT13946) in
CC plasmid pCIB5531. VIP2A(a) acts as an auxiliary protein to
CC VIP1A(a), an insect-specific protein showing activity against
CC western corn rootworm. The construct may be expressed in
CC transgenic plants, esp. maize.
XX
SQ Sequence 1338 AA;

Query Match 6.7%; Score 112.5; DB 17; Length 1338;
Best Local Similarity 23.5%; Pred. No. 0.39;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

```
Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       |||| |:  |    || ||: |: ::      | : :
Db     568 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 620

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     621 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqqmkr 674

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     675 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 732

Qy     179 ARAHK-MRQILMNNGYD---------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :                ||  :  |   |           | :
Db     733 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 787

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
             || |  | : :  | |   | |              |  :  |||  : : : | |
Db     788 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 847

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     848 rynnvgtgaiydvkpttsfvlnndtiat 875
```

RESULT 15
AAW19520
ID AAW19520 standard; Protein; 1338 AA.
XX
AC AAW19520;
XX
DT 15-APR-1998 (first entry)
XX
DE Maize optimised-B. cereus VIP2A(a)/VIP1A(a) fusion protein.
XX
KW Vegetative insecticidal protein; Bacillus cereus strain AB78; plant;
KW insect; Sesamia nonagrioides; maize; corn borer; toxin; fusion protein.
XX
OS Bacillus cereus.
OS Synthetic.
XX
PN WO9726339-A1.
XX
PD 24-JUL-1997.
XX
PF 23-DEC-1996; 96WO-EP05828.
XX
PR 15-JAN-1996; 96GB-0000786.
XX
PA (NOVS) NOVARTIS AG.
XX
PI Gay PB;
XX
DR WPI; 1997-385342/35.
DR N-PSDB; AAT74010.
XX
PT Protecting plants against insects of the genus Sesamia using
PT Bacillus toxic proteins - applied directly or expressed as
PT heterologous protein by the plant, also transgenic plants expressing
PT both Cry and VIP type toxins
XX
PS Claim 8; Page 124-130; 168pp; English.
XX
CC This is the amino acid sequence of a fusion protein comprising the
CC Bacillus cereus vegetative insecticidal protein (VIP) 2A(a) and VIP1A(a)
CC which have been optimised for expression in maize plants. The VIP
CC fusion protein can be used in a new method for protecting plants, and
CC their progeny, against insects of the genus Sesamia by direct or indirect
CC application to the plant (or seed or growing area). The protein is

```
CC   especially useful to protect maize plants against the Mediterranean
CC   corn borer (S. nonagrioides).
XX
SQ   Sequence   1338 AA;


  Query Match             6.7%;  Score 112.5;  DB 18;  Length 1338;
  Best Local Similarity   23.5%;  Pred. No. 0.39;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     568 tfnlsedeqaiieingkiisnkgkekqvvhlek----gklvpikieyqs---dtkfnids 620

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     621 ktfkelklfkidsq--nqpqqvqqdelrnpefnkkesqef----lakpskinlftqqmkr 674

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     675 eidedtdtdgdsipdlweengytiqnriavkwddslaskgytkfvsnples--htvgdpy 732

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               |  | : ::|  :               ||  :  |   |          | :
Db     733 tdyekaardldlsnaketfnplvaafpsvnvsmekvilspnenlsnsv-----eshsstn 787

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
            || |  | : :  | |   | |            |  :  |||  : : : | |
Db     788 wsytntegasveagigpkgisfgvsvnyqhsetvaqewgtstgntsqfntasagylnanv 847

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     848 rynnvgtgaiydvkpttsfvlnndtiat 875


Search completed: June 27, 2002, 21:47:24
Job time: 5512 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:48:14 ; Search time 38.03 Seconds
(without alignments)
205.527 Million cell updates/sec

Title: US-09-727-769A-8
Perfect score: 1679
Sequence: 1 MKNLFLSMMAFVTVLTFNSC.....FSLLSGCSPSPAPDVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1292.5	77.0	319	4	US-09-324-910-12	Sequence 12, Appl
2	874	52.1	185	4	US-09-324-910-6	Sequence 6, Appli
3	112.5	6.7	852	1	US-08-471-033-36	Sequence 36, Appl
4	112.5	6.7	852	2	US-08-471-044-36	Sequence 36, Appl
5	112.5	6.7	852	2	US-08-463-483A-36	Sequence 36, Appl
6	112.5	6.7	852	2	US-08-471-046A-36	Sequence 36, Appl
7	112.5	6.7	852	2	US-08-470-566B-36	Sequence 36, Appl
8	112.5	6.7	852	2	US-08-469-334-36	Sequence 36, Appl
9	112.5	6.7	852	3	US-09-300-529-36	Sequence 36, Appl
10	112.5	6.7	881	4	US-08-960-780-32	Sequence 32, Appl
11	112.5	6.7	881	4	US-09-073-898-32	Sequence 32, Appl
12	112.5	6.7	884	1	US-08-471-033-5	Sequence 5, Appli
13	112.5	6.7	884	2	US-08-471-044-5	Sequence 5, Appli
14	112.5	6.7	884	2	US-08-463-483A-5	Sequence 5, Appli
15	112.5	6.7	884	2	US-08-471-046A-5	Sequence 5, Appli
16	112.5	6.7	884	2	US-08-470-566B-5	Sequence 5, Appli
17	112.5	6.7	884	2	US-08-469-334-5	Sequence 5, Appli
18	112.5	6.7	884	3	US-09-300-529-5	Sequence 5, Appli
19	112.5	6.7	1338	1	US-08-471-033-50	Sequence 50, Appl
20	112.5	6.7	1338	2	US-08-471-044-50	Sequence 50, Appl
21	112.5	6.7	1338	2	US-08-463-483A-50	Sequence 50, Appl
22	112.5	6.7	1338	2	US-08-471-046A-50	Sequence 50, Appl
23	112.5	6.7	1338	2	US-08-470-566B-50	Sequence 50, Appl
24	112.5	6.7	1338	2	US-08-469-334-50	Sequence 50, Appl
25	112.5	6.7	1338	3	US-09-300-529-50	Sequence 50, Appl
26	112.5	6.7	1346	1	US-08-471-033-23	Sequence 23, Appl
27	112.5	6.7	1346	2	US-08-471-044-23	Sequence 23, Appl
28	112.5	6.7	1346	2	US-08-463-483A-23	Sequence 23, Appl
29	112.5	6.7	1346	2	US-08-471-046A-23	Sequence 23, Appl
30	112.5	6.7	1346	2	US-08-470-566B-23	Sequence 23, Appl
31	112.5	6.7	1346	2	US-08-469-334-23	Sequence 23, Appl
32	112.5	6.7	1346	3	US-09-300-529-23	Sequence 23, Appl
33	95	5.7	1541	4	US-08-296-791-3	Sequence 3, Appli
34	95	5.7	1541	5	PCT-US95-10661A-3	Sequence 3, Appli
35	93	5.5	1183	2	US-08-447-031A-2	Sequence 2, Appli
36	91.5	5.4	615	2	US-08-525-742-10	Sequence 10, Appl
37	90.5	5.4	1222	2	US-08-682-517-15	Sequence 15, Appl
38	90.5	5.4	1252	2	US-08-682-517-9	Sequence 9, Appli
39	89.5	5.3	610	2	US-08-525-742-8	Sequence 8, Appli
40	89.5	5.3	1338	2	US-08-728-470-9	Sequence 9, Appli
41	89.5	5.3	1338	4	US-08-719-641-9	Sequence 9, Appli
42	89.5	5.3	1599	2	US-08-617-697-9	Sequence 9, Appli
43	88.5	5.3	512	4	US-08-856-253-6	Sequence 6, Appli
44	88	5.2	834	1	US-08-471-033-21	Sequence 21, Appl
45	88	5.2	834	2	US-08-471-044-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-324-910-12
; Sequence 12, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-09-324-910-12

Query Match 77.0%; Score 1292.5; DB 4; Length 319;
Best Local Similarity 74.7%; Pred. No. 5.5e-129;
Matches 239; Conservative 39; Mismatches 41; Indels 1; Gaps 1;

```
Qy       1 MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMI 60
           ||   ||||||||:|:||:|:||: ||: |   | | |:  :||||||||||::|:|
Db       1 MKKFLLSMMAFVTILSFNACSDSSANQDPNLVAKES-NEVAMKDFGKTVPVGIEKEDGKF 59

Qy      61 KVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRY 120
           |:|||:||| |||  :|||  || |:||||:||:|||:||| |:|:| ||| |: :|:||
Db      60 KISFMVTAQPYEIADSKENAGYISMIRQAVENETPVHVFLKVNTNKIAKVEKATDDDIRY 119

Qy     121 FKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYAR 180
           ||::  |: :|::||  |||||:||||||| |||||:|||||||||||||||||||||||
Db     120 FKSVFNKQERGESNKAVSVIPDLATLNSLFTQIKNQACGTSTASSPCITFRYPVDGCYAR 179

Qy     181 AHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPS 240
           ||||||||:| ||||||||||||:||||||||:| |||||||:|||||: ||||||||
Db     180 AHKMRQILLNAGYDCEKQFVYGNLRASTGTCCVSWVYHVAILVSFKNASGIVEKRIIDPS 239

Qy     241 LFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTK 300
           |||||||||:||| || |||||||||||||||||||||||| | ||||| :||||||
Db     240 LFSSGPVTDSAWRAACTNTSCGSASVSSYANTAGNVYYRSPSGSLLYDNNYVNTNCVLNI 299

Qy     301 FSLLSGCSPSPAPDVSSCGF 320
           || |||||||||| |:||||
Db     300 FSSLSGCSPSPAPSVASCGF 319
```

```
RESULT   2
US-09-324-910-6
; Sequence 6, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
;  APPLICANT: Yamaguchi, Shotaro
;  APPLICANT: Matsuura, Akira
;  TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
;  TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
;  FILE REFERENCE: Q-54362
;  CURRENT APPLICATION NUMBER: US/09/324,910
;  CURRENT FILING DATE: 1999-06-03
;  EARLIER APPLICATION NUMBER: HEI-10-173940
;  EARLIER FILING DATE: 1998-06-04
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 185
;   TYPE: PRT
;   ORGANISM: Chryseobacterium gleum
US-09-324-910-6


  Query Match             52.1%;  Score 874;  DB 4;  Length 185;
  Best Local Similarity   87.4%;  Pred. No. 6.8e-85;
  Matches  160;  Conservative   10;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     138 SVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEK 197
           |||||:||||||| ||||||:|||||||||||||||||||||||||||||||:| ||||||
Db       3 SVIPDLATLNSLFTQIKNQACGTSTASSPCITFRYPVDGCYARAHKMRQILLNAGYDCEK 62

Qy     198 QFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACV 257
           |||||||:||||||||:| |||||||||:|||||: ||||||||||||||||:||| ||
Db      63 QFVYGNLRASTGTCCVSWVYHVAILVSFKNASGIVEKRIIDPSLFSSGPVTDSAWRAACT 122

Qy     258 NTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSS 317
           |||||||||||||||||||||||| | ||||| :||||||  || ||||||||||| |:|
Db     123 NTSCGSASVSSYANTAGNVYYRSPSGSLLYDNNYVNTNCVLNIFSSLSGCSPSPAPSVAS 182

Qy     318 CGF 320
           |||
Db     183 CGF 185


RESULT   3
US-08-471-033-36
; Sequence 36, Application US/08471033
; Patent No. 5770696
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  852 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-471-033-36


  Query Match              6.7%;  Score 112.5;  DB 1;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.0076;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       |||| |:  |    || ||: |: ::      | : :
Db      82 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     135 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQQMKR 188

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:  |:    |||:   |  :: |:|  :      | | |   |:|  :  : | |
Db     189 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 246

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               |  | : ::|  :                 ||  :  |   |          | :
Db     247 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 301

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
            || |  | : :  | |   | |              |  :  |||  : : : | |
Db     302 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 361

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     362 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 389


RESULT   4
US-08-471-044-36
; Sequence 36, Application US/08471044
; Patent No. 5840868
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
```

```
;    TITLE OF INVENTION:  No. 5840868el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,044
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV6 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  852 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-471-044-36


  Query Match             6.7%;  Score 112.5;  DB 2;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.0076;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db      82 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     135 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQQMKR 188

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     189 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 246

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :              || :  |   |          | :
Db     247 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 301

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |               |  :  |||  : : : | |
Db     302 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 361

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                  | :|   |: |:: :|: | |
Db     362 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 389
```

```
RESULT   5
US-08-463-483A-36
; Sequence 36, Application US/08463483A
; Patent No. 5849870
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5849870el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,483A
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  852 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-463-483A-36


  Query Match             6.7%;  Score 112.5;  DB 2;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.0076;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db      82 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     135 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQQMKR 188

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     189 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 246

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :              || :  |   |          | :
Db     247 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 301

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |               |  :  |||  : : : | |
Db     302 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 361

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                  | :|   |: |:: :|: | |
Db     362 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 389
```

```
Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db      82 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     135 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQQMKR 188

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
```

```
Db    189 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 246

Qy    179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
Db    247 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 301

Qy    223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
Db    302 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 361

Qy    273 ------AGNVYYRSPSNSYLYDNNLINT 294
Db    362 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 389


RESULT   6
US-08-471-046A-36
; Sequence 36, Application US/08471046A
; Patent No. 5866326
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Method For Isolating Vegetative Insecticidal
;    TITLE OF INVENTION:  Protein Genes
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5866326artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,046A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV8 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  852 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein

US-08-471-046A-36


  Query Match             6.7%;  Score 112.5;  DB 2;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.0076;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy     16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
Db     82 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 134

Qy     69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
Db    135 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQQMKR 188

Qy    128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
Db    189 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 246

Qy    179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
Db    247 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 301

Qy    223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
Db    302 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 361

Qy    273 ------AGNVYYRSPSNSYLYDNNLINT 294
Db    362 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 389


RESULT   7
US-08-470-566B-36
; Sequence 36, Application US/08470566B
; Patent No. 5872212
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5872212el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5872212artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,566B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV4 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  852 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-470-566B-36


  Query Match             6.7%;  Score 112.5;  DE 2;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.0076;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db      82 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     135 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQQMKR 188

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     189 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 246

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
           |  | : ::|  :                 ||  :  |   |          | :
Db     247 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 301

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |              |  :  |||  : : : | |
Db     302 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 361

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     362 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 389


RESULT   8
US-08-469-334-36
; Sequence 36, Application US/08469334
; Patent No. 5990383
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5990383el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,334
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/463,483
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  852 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-469-334-36


  Query Match             6.7%;  Score 112.5;  DB 2;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.0076;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db      82 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     135 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQQMKR 188

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     189 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 246

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
           |  | : ::|  :                 ||  :  |   |          | :
Db     247 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 301

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |              |  :  |||  : : : | |
Db     302 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 361

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     362 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 389


RESULT   9
US-09-300-529-36
; Sequence 36, Application US/09300529
; Patent No. 6066783
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
```

```
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Genes Encoding Insecticidal Proteins
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6066783artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/300,529
;      FILING DATE:  TBA
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/469,334
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  S-19506L
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  852 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-300-529-36


  Query Match             6.7%;  Score 112.5;  DB 3;  Length 852;
  Best Local Similarity   23.5%;  Pred. No. 0.0076;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::       | : :
Db      82 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 134

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     135 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQQMKR 188

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : | |
Db     189 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 246

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :                ||  :  |   |        | :
Db     247 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 301

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |              |  :  |||  : : : | |
Db     302 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 361

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     362 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 389


RESULT  10
US-08-960-780-32
; Sequence 32, Application US/08960780
; Patent No. 6204435
;  GENERAL INFORMATION:
;    APPLICANT:  Feitelson, Jerald S.
;    APPLICANT:  Schnepf, H. Ernest
;    APPLICANT:  Narva, Kenneth E.
;    APPLICANT:  Stockhoff, Brian A.
;    APPLICANT:  Schmeits, James
;    APPLICANT:  Loewer, David
;    APPLICANT:  Dullum, Charles Joseph
;    APPLICANT:  Muller-Cohn, Judy
;    APPLICANT:  Stamp, Lisa
;    TITLE OF INVENTION:  No. 6204435el Pesticidal Toxins and Nucleotide
;    TITLE OF INVENTION:  Sequences Which Encode These Toxins
;    NUMBER OF SEQUENCES:  134
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  US
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/960,780
;      FILING DATE:  30-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/029,848
;      FILING DATE:  30-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  MA-708
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  352-375-8100
;      TELEFAX:  352-372-5800
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  881 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  177C8
US-08-960-780-32


  Query Match             6.7%;  Score 112.5;  DB 4;  Length 881;
  Best Local Similarity   23.5%;  Pred. No. 0.008;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::       | : :
Db     111 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 163

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
```

```
          : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db    164 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQKMKR 217

Qy    128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
          |:   |:     |||:   |  :: |:|  :       | | |   |:|  :  : |   |
Db    218 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 275

Qy    179 ARAHK-MRQILMNNGYD---------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
              |  | : ::|  :                 ||  :  |   |          | :
Db    276 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 330

Qy    223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |           |  :  |||  : : : | |
Db    331 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 390

Qy    273 ------AGNVYYRSPSNSYLYDNNLINT 294
                | :|   |: |:: :|: | |
Db    391 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 418


RESULT  11
US-09-073-898-32
; Sequence 32, Application US/09073898
; Patent No. 6242669
;  GENERAL INFORMATION:
;    APPLICANT:  Feitelson, Jerald S.
;    APPLICANT:  Schnepf, H. Ernest
;    APPLICANT:  Narva, Kenneth E.
;    APPLICANT:  Stockhoff, Brian A.
;    APPLICANT:  Schmeits, James
;    APPLICANT:  Loewer, David
;    APPLICANT:  Dullum, Charles Joseph
;    APPLICANT:  Muller-Cohn, Judy
;    APPLICANT:  Stamp, Lisa
;    APPLICANT:  Morrill, George
;    APPLICANT:  Finstad-Lee, Stacey
;    TITLE OF INVENTION:  No. 6242669el Pesticidal Toxins and Nucleotide
;    TITLE OF INVENTION:  Sequences Which Encode These Toxins
;    NUMBER OF SEQUENCES:  144
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  US
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,898
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/029,848
;      FILING DATE:  30-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/960,780
;      FILING DATE:  30-OCT-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sanders, Jay M.
;      REGISTRATION NUMBER:  39,355
;      REFERENCE/DOCKET NUMBER:  MA-708C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  352-375-8100
;      TELEFAX:  352-372-5800
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  881 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  PS177C8
US-09-073-898-32


  Query Match             6.7%;  Score 112.5;  DB 4;  Length 881;
  Best Local Similarity   23.5%;  Pred. No. 0.008;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy     16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
          |||   |     |||       ||||  |:  |    || ||: |: ::       | : :
Db    111 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 163

Qy     69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
          : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db    164 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQKMKR 217

Qy    128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
          |:   |:     |||:   |  :: |:|  :       | | |   |:|  :  : |   |
Db    218 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 275

Qy    179 ARAHK-MRQILMNNGYD---------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
              |  | : ::|  :                 ||  :  |   |          | :
Db    276 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 330

Qy    223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |           |  :  |||  : : : | |
Db    331 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 390

Qy    273 ------AGNVYYRSPSNSYLYDNNLINT 294
                | :|   |: |:: :|: | |
Db    391 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 418


RESULT  12
US-08-471-033-5
; Sequence 5, Application US/08471033
; Patent No. 5770696
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
```

```
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  884 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-471-033-5


  Query Match             6.7%;  Score 112.5;  DB 1;  Length 884;
  Best Local Similarity   23.5%;  Pred. No. 0.0081;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     114 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 166

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     167 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQKMKR 220

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     221 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 278

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :                ||  :  |   |          | :
Db     279 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 333

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |               |  :  |||  : : : | |
Db     334 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 393

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     394 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 421


RESULT  13
US-08-471-044-5
; Sequence 5, Application US/08471044
; Patent No. 5840868
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J .
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5840868el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,044
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV6 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  884 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-471-044-5


  Query Match             6.7%;  Score 112.5;  DB 2;  Length 884;
  Best Local Similarity   23.5%;  Pred. No. 0.0081;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     114 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 166

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     167 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQKMKR 220

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     221 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 278

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               | | : ::|  :                ||  :  |   |          | :
Db     279 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 333

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |               |  :  |||  : : : | |
Db     334 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 393

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                 | :|   |: |:: :|: | |
Db     394 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 421


RESULT  14
US-08-463-483A-5
; Sequence 5, Application US/08463483A
; Patent No. 5849870
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
```

```
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5849870el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,483A
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  884 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-463-483A-5


  Query Match             6.7%;  Score 112.5;  DB 2;  Length 884;
  Best Local Similarity   23.5%;  Pred. No. 0.0081;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     114 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 166

Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     167 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQKMKR 220

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     |||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     221 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 278

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
             |  | : ::|  :                ||  :  |   |           | :
Db     279 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 333

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
           || |  | : :  | |   | |               |  :  |||  : : : | |
Db     334 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 393

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                  | :|   |: |:: :|: | |
Db     394 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 421


RESULT  15
US-08-471-046A-5
; Sequence 5, Application US/08471046A
; Patent No. 5866326
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Method For Isolating Vegetative Insecticidal
;    TITLE OF INVENTION:  Protein Genes
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5866326artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,046A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV8 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  884 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-471-046A-5


  Query Match             6.7%;  Score 112.5;  DB 2;  Length 884;
  Best Local Similarity   23.5%;  Pred. No. 0.0081;
  Matches   77;  Conservative   44;  Mismatches  138;  Indels   69;  Gaps   15;

Qy      16 TFNSCADSNGNQEIN-------GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTA 68
           |||   |     |||       ||||  |:  |    || ||: |: ::      | : :
Db     114 TFNLSEDEQAIIEINGKIISNKGKEKQVVHLEK----GKLVPIKIEYQS---DTKFNIDS 166
```

```
Qy      69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK 127
           : | |:|  | : |     :|  ::|     | |  | |      | |  :  |   : :
Db     167 KTFKELKLFKIDSQ--NQPQQVQQDELRNPEFNKKESQEF----LAKPSKINLFTQKMKR 220

Qy     128 EVKGQTNKLASVIPDVATLN--SLFNQIKNQ------SCG-TSTASSPCITFRYPVDGCY 178
           |:   |:     ||:   |  :: |:|  :      | | |   |:|  :  : |   |
Db     221 EIDEDTDTDGDSIPDLWEENGYTIQNRIAVKWDDSLASKGYTKFVSNPLES--HTVGDPY 278

Qy     179 ARAHK-MRQILMNNGYD--------------CEKQFVYGNLKASTGTCCVAWSYHVAIL 222
               |  | : ::|  :                ||  :  |   |           | :
Db     279 TDYEKAARDLDLSNAKETFNPLVAAFPSVNVSMEKVILSPNENLSNSV-----ESHSSTN 333

Qy     223 VSYKNASGVTEKRIIDPSLFSSG--------PVTDTAWRNACVNTSCGSASVSSYANT-- 272
            || |  | : :  | |   | |              |  :  |||  : : : | |
Db     334 WSYTNTEGASVEAGIGPKGISFGVSVNYQHSETVAQEWGTSTGNTSQFNTASAGYLNANV 393

Qy     273 ------AGNVYYRSPSNSYLYDNNLINT 294
                | :|   |: |:: :|: | |
Db     394 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 421
```

Search completed: June 27, 2002, 21:48:16
Job time: 5364 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:49:25 ; Search time 55.76 Seconds
(without alignments)
551.445 Million cell updates/sec

Title: US-09-727-769A-8
Perfect score: 1679
Sequence: 1 MKNLFLSMMAFVTVLTFNSC.....FSLLSGCSPSPAPDVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	108	6.4	927	2	T21772	hypothetical prote
2	101.5	6.0	551	2	A83232	conserved hypothet
3	101	6.0	2395	1	S50820	surface protein ty
4	99.5	5.9	517	1	GIBPT4	gene 12 protein -
5	99.5	5.9	534	2	T39903	serine-rich protei
6	99	5.9	298	2	I64138	adhesin homolog HI
7	99	5.9	564	2	JQ1792	SalF17R protein -
8	99	5.9	594	2	T50765	adhesion of calyx
9	98.5	5.9	419	2	C83912	hypothetical prote
10	97	5.8	534	2	C69146	hypothetical prote
11	96.5	5.7	260	2	B90026	hypothetical prote
12	96.5	5.7	490	2	H70103	hypothetical prote
13	95.5	5.7	1679	2	S49802	probable membrane
14	95	5.7	564	2	C42523	A55R protein - vac
15	95	5.7	1541	2	A37023	IgA-specific metal
16	94.5	5.6	622	2	A64494	hypothetical prote
17	93.5	5.6	1894	2	JC4980	plexin 1 precursor
18	93	5.5	594	2	T50764	adhesion of calyx
19	92.5	5.5	1582	2	AC1153	adhesin homolog lm
20	92	5.5	1154	2	A86318	protein F15H18.11
21	91.5	5.4	627	2	B64710	site-specific DNA-
22	91.5	5.4	851	2	S67285	NUD1 protein - yea
23	91.5	5.4	2013	2	AI1489	probable peptidogl
24	91	5.4	396	2	T52596	squamosa promoter
25	91	5.4	761	1	IJHUNG	neural cell adhesi
26	90.5	5.4	843	2	T22182	hypothetical prote
27	90	5.4	243	2	T28802	hypothetical prote
28	89.5	5.3	1578	2	AD1512	peptidoglycan boun
29	89.5	5.3	2500	2	G71609	hypothetical prote
30	89.5	5.3	2510	2	T28160	hypothetical prote
31	89	5.3	595	2	B86798	prophage pi3 prote
32	89	5.3	1073	2	D87596	TonB-dependent rec
33	89	5.3	1895	2	T15881	hypothetical prote
34	88.5	5.3	283	2	G69212	conserved hypothet
35	88.5	5.3	466	2	H96991	secreted protein c
36	88.5	5.3	491	2	T20061	hypothetical prote
37	88.5	5.3	596	2	AC1318	two-component sens
38	88.5	5.3	685	2	AF0850	hypothetical prote
39	88.5	5.3	693	2	T15728	hypothetical prote
40	88.5	5.3	1251	2	T21389	hypothetical prote
41	88.5	5.3	1341	2	S66835	probable membrane
42	88	5.2	603	2	AE2119	GTP-binding elonga
43	88	5.2	814	2	T02195	hypothetical prote
44	88	5.2	2013	2	AD1129	probable peptidogl
45	87.5	5.2	372	2	F86189	hypothetical prote

ALIGNMENTS

RESULT 1
T21772
hypothetical protein T01D3.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T21772; T24296
R;Steward, C.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19470
A;Accession: T21772
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-927 <WIL>
A;Cross-references: EMBL:Z81527; PIDN:CAB04279.1; GSPDB:GN00023; CESP:T01D3.6
A;Experimental source: clone F35E12
R;Steward, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19870
A;Accession: T24296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-927 <WI2>
A;Cross-references: EMBL:Z81110; PIDN:CAB03262.1; GSPDB:GN00023; CESP:T01D3.6
A;Experimental source: clone T01D3
C;Genetics:
A;Gene: CESP:T01D3.6
A;Map position: 5
A;Introns: 38/2; 87/3; 128/2; 242/2; 374/3; 416/3; 734/3; 842/3; 857/1

Query Match 6.4%; Score 108; DB 2; Length 927;
Best Local Similarity 19.2%; Pred. No. 2.8;
Matches 85; Conservative 47; Mismatches 101; Indels 210; Gaps 24;

```
Qy      35 LSVNDSKLKDFGKTVPV----------GIDEENGMIKVSFMLTAQFYEIKPTKENEQYI 83
           |:  :  || :|    |           |:: |:   | : :     |:    :::::|:
Db      15 LAAAEDDLKCYGADCEVPHYMDDTEQCSGLECES---KETLLAMISHYQ---KEKHDEYL 68

Qy      84 GMLRQAVKNESPVHIFLKPNSNEIG----KVESASPE----------------DVR--- 119
             | : :::|| |  ||    ||:       : ||:                 | |
Db      69 APLTKPIRDESAVRNFL----NEVAVTGDLAATVSPQCSICSQPGLCNSGQCVPDARFPW 124

Qy     120 -YFKTILTKEVKGQTNKLASVIPDVATLNSLFNQI--KNQSCGTSTASSPCITFRYPVDG 176
            ||                | || |:     |:|  |: |||  :
Db     125 QYF--------------YCVCPDYASGRFCQNEIKCKDNSCGKNA----------D 156

Qy     177 CYARAHKMRQI------LMNNGYDCEKQ-----------FVYGNLKAS-TGTCCVAWS- 216
           ||   |::  |         || ||: :             |  |:    |||   :|
Db     157 CYVANHQLNCICKPGYTARRNGRDCDMKVQQACMSGDPHYVTYDGLRFDYQGTCPYVFSQ 216

Qy     217 ----------------------YHVA----ILVSYKN----------ASGVTEKR 235
                                ||::    : |   |             :||   :
```

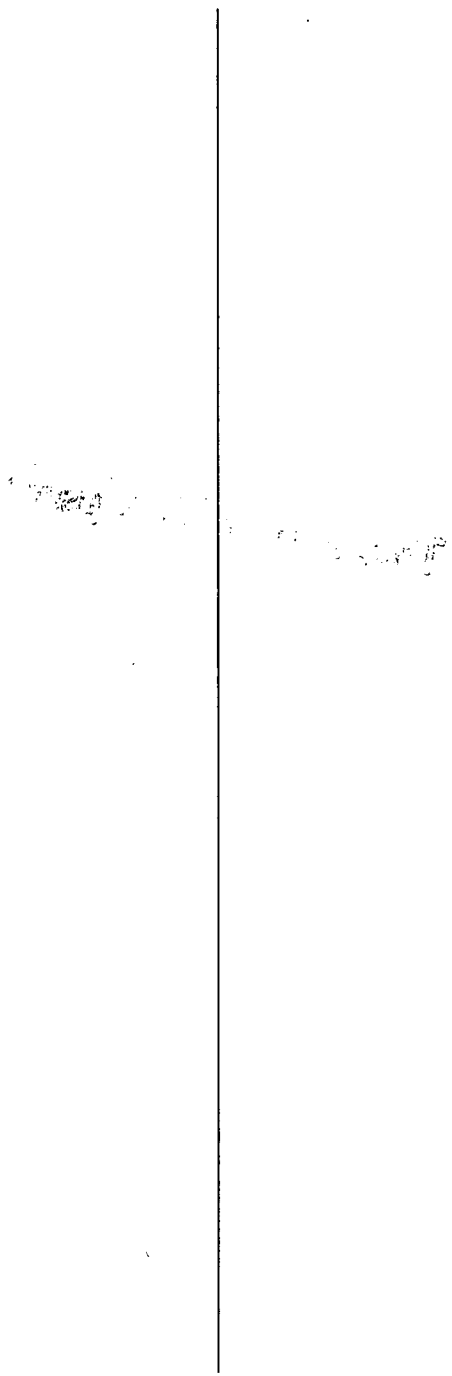
```
Db    217 PCTTLPAPYLWYSVRAKNELPGKGYHISQVSEVEVDLHNLTIHVDGRSKTALVNGV---Q 273

Qy    236 IIDPSLFS-----------SGP-----------VTDTAWRNACV-----------NTSCG 262
          :: |  |            ||            || | : : ||           | ||
Db    274 VLTPWYFPNKNTWTVRVRFSGSTFTIENDQGVVVTFTTYNSLCVQVPDIPEFNGATTLCG 333

Qy    263 SA------SVSSYANTAGNVY----YRSPSN-----------SYLYDNNLI--------- 292
            |       :    |  |:|      | | |           ::: |  ||
Db    334 LAGNIDGKKLDDVVNKNGSVLAIKSSRQPENNNHADFMKTEDTWITDKFLILRPGQENCI 393

Qy    293 -------NTNCVLTKFSLLSGCS 308
                 ||||| |  ||   |:
Db    394 NGQTLDNNTNCVSTSISLAQSCA 416


RESULT   2
A83232
conserved hypothetical protein PA3310 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83232
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: A83232
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-551 <STO>
A;Cross-references: GB:AE004753; GB:AE004091; NID:g9949433; PIDN:AAG06698.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3310
C;Superfamily: Escherichia coli yidB protein


  Query Match             6.0%;  Score 101.5;  DB 2;  Length 551;
  Best Local Similarity   20.1%;  Pred. No. 4.4;
  Matches   64;  Conservative   40;  Mismatches   95;  Indels  119;  Gaps   15;

Qy     58 GMIKVSFMLTAQFYE---------------IKPTKENEQYIGMLRQAVKNESPV--HIFL 100
          | : :  ::   ||                | ||     ||  : :  |    :  |: :
Db    163 GCLLIVAVMVGSFYSTYAPIFREEDKLTHFINPT----NYIYAVGKYAKQRLVIKEHLVV 218

Qy    101 KPNSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNSL---------- 149
          :|    ||     :|  ::  |  |   | |:| :      |  :||
Db    219 EP----IGLDARQAPAALQRQKKSLLVFVVGETAR-----ADHFSLNGYARETNPELSKQ 269

Qy    150 ----FNQIKNQSCGTSTA-SSPCITFRYPVDGC---YARAHK----------MRQILMNN 191
              | |::  ||||||| | ||:  :|| :      |: |:          :: : : |
Db    270 DIVNFTQVR--SCGTSTAVSVPCMFSQYPREDYSDKKAKTHEGLLDILQRAGVQVLWLEN 327

Qy    192 GYDCEKQFVYGNLKASTGTC--------------------CVAWSYHVAILVSYKNASG 230
            ||:            |||                    |:  |  | :    :   |
Db    328 NSDCK-----------GTCLRVPNRDIPKTQPSPFCDGKNCLDESLLVGL---QEYIDG 372

Qy    231 VTEKRII--------DPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPS 282
          : :  ||         |  :   |     ::  |     || |     |
Db    373 LQDDAIIVLHSDGSHGPEYYERYPKEMERFQPVCRTNQLGSCSKEELVN---------- 421

Qy    283 NSYLYDNNLINTNCVLTK 300
             :||| :: |:  |||
Db    422 ---VYDNTILYTDHFLTK 436


RESULT   3
S50820
surface protein type 51B - Paramecium tetraurelia
C;Species: Paramecium tetraurelia
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
C;Accession: S50820
R;Scott, J.; Leeck, C.; Forney, J.
Nucleic Acids Res. 22, 5079-5084, 1994
A;Title: Analysis of the micronuclear B type surface protein gene in Paramecium tetra
A;Reference number: S50820; MUID:95098630
A;Accession: S50820
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2395 <SCO>
A;Cross-references: EMBL:U07603; NID:g467226; PIDN:AAA81947.1; PID:g467227
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C;Genetics:
A;Genetic code: SGC5
A;Introns: 472/3; 1310/3; 1821/3
C;Superfamily: G surface protein


  Query Match             6.0%;  Score 101;  DB 1;  Length 2395;
  Best Local Similarity   22.3%;  Pred. No. 36;
  Matches   62;  Conservative   22;  Mismatches   94;  Indels  100;  Gaps   12;

Qy    122 KTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVD------ 175
          ||          || |         |:||       :::|  | | : |       |
Db    464 KTCANAPSTNTTNDLCVAFLSSCTVNSTNAGCVDKTCENSLAQTIC-------DKDLNNK 516

Qy    176 ------GCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCV-------AWSYHVAIL 222
                 || |    :          ||:    | |   ::||| ||       | :   |
Db    517 ACIWKGKCYKRECVLASSTTATHADCQTYDV-GCTLSNTGTGCVPLPLKCEAITIEAACN 575

Qy    223 VSYKNASGV----------------------TEKRIIDPSLFSSG-------------- 245
          :  :  |||                      |      |   : ||
Db    576 IRLQVTSGVKSYPLCGWNGSSCIDKACSTAPKTTATTSDCGTYKSGCVANNPVNGSIQGC 635

Qy    246 ---PVTDTA-----------------WR---NACVNTSCGSASVSSYANTAGNVYYRSP 281
             | |  |                 |    :|||  || :|||::   | | :   |
Db    636 QDLPTTCAARKSTENCEITRTGFPTCLWNSATSACVEKSCSTASVTT---TTGFLTVFSN 692

Qy    282 SNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCG 319
          :|   |   | |: |:    :   || | |    ||||
Db    693 TNCLAY---LSNSACIAN--NTADGCIPKP----SSCG 721


RESULT   4
GIBPT4
gene 12 protein - phage T4
N;Alternate names: tail fiber protein
C;Species: phage T4
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 30-Jun-1993
C;Accession: S01889
R;Selivanov, N.A.; Prilipov, A.G.; Mesyanzhinov, V.V.
Nucleic Acids Res. 16, 2334, 1988
A;Title: Nucleotide and deduced amino acid sequence of bacteriophage T4 gene 12.
A;Reference number: S01889; MUID:88189824
A;Accession: S01889
A;Molecule type: DNA
A;Residues: 1-517 <SEL>
A;Cross-references: EMBL:X06792
A;Note: the authors translated the codon CAG for residue 279 as His
C;Genetics:
A;Gene: 12
C;Superfamily: phage T4 gene 12 protein
C;Keywords: tail fiber


  Query Match             5.9%;  Score 99.5;  DB 1;  Length 517;
  Best Local Similarity   20.8%;  Pred. No. 5.8;
  Matches   77;  Conservative   45;  Mismatches  134;  Indels  115;  Gaps   15;

Qy     24 NGNQEINGKEKLSVNDSKLKDFGK---------TVP--------VGIDEENGMIKVSFML 66
          |  : : |  : | ||  :               ||         |  :  ||:||:| :
Db     79 NATETVYGLTRYSTNDEAIAGVNNESSITPAKFTVALNNAFETRVSTESSNGVIKISSLP 138

Qy     67 TA-----QFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVR-- 119
           |        : | |  :  | :: |   :|:          :: | |: |:   ||
```

```
Db    139 QALAGADDTTAMTPLKTQQLAIKLIAQIAPSETTA------TESDQGVVQLATVAQVRQG 192

Qy    120 ------------YFKTILTKEVKG---------QTNKLASVIPDVATLN------SLFNQ 152
                      :  :  |:| ||          :  |||    ||||      |:
Db    193 TLREGYAISPYTFMNSSSTEEYKGVIKLGTQSEVNSNNASVAVTGATLNGRGSTTSMRGV 252

Qy    153 IKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKAS----- 207
          :|     |:|| |     :   |   | |     |    |     | :|| |:
Db    253 VKL----TTTAGS-----QSGGDASSALAWNADVIQQRGG-----QIIYGTLRIEDTFTI 298

Qy    208 -------TGTCCVAWSYHVAILVSYKNASGVTEKRIIDPS-------LFSSGPVTDTAWR 253
                |||        | :   |   : :  :  || :       ::::  :   |||
Db    299 ANGGANITGT--------VRMTGGYIQGNRIVTQNEIDRTIPVGAIMMWAADSLPSDAWR 350

Qy    254 NACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYD----------------NNLINTNCV 297
            |   :  ::    ||:  |  |  :|||  | |                 |  :| |
Db    351 -FCHGGTVSASDCPLYASRIGTRYGGNPSNPGLPDMRGLFVRGSGRGSHLTNPNVNGNDQ 409

Qy    298 LTKFSLLSGCS 308
            |  |  ||:
Db    410 FGKPRLGVGCT 420


RESULT   5
T39903
serine-rich protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C;Accession: T39903
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z21889
A;Accession: T39903
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-534 <LYN>
A;Cross-references: EMBL:AL033534; PIDN:CAA22127.1; GSPDB:GN00067; SPDB:SPBC215.13
A;Experimental source: strain 972h-; cosmid c215
C;Genetics:
A;Gene: SPDB:SPBC215.13
A;Map position: 2
C;Superfamily: pig submaxillary mucin


  Query Match             5.9%;  Score 99.5;  DB 2;  Length 534;
  Best Local Similarity   22.7%;  Pred. No. 6.1;
  Matches   55;  Conservative   39;  Mismatches  123;  Indels   25;  Gaps    6;

Qy     85 MLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVA 144
          :: |   | |   | |        : |    :::||:      |  |         |  :
Db     24 LVHQIYGNSSFTKISLNQLEGRDSQEELQRRQEIRYYGR--AAETGGTPTYYGYATPTSS 81

Qy    145 TLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNL 204
          :  |:|::    |  |:: |||  ::  |            |:  :  :       |
Db     82 SEPSLFSESATPS-ETNSYSSPVSSYSDPATS---------QLPSSTSFFSPTSSEYTPS 131

Qy    205 KASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSA 264
          : :      |   ||| |    |   |  |   || || |:| : : :   :||
Db    132 STESSSLLDPSSVSSAILPS----STSVEVSISSSSLSSSDPLTSSTFSSLSSSTSSSQP 187

Qy    265 SVSS-YANTAGNVYYRSPSNSYLYDNNLI--------NTNCVLTKFSLLSGCSPSPAPDV 315
          ||||  ::|  :    | |:|||  ::::        :::  ||  || :   || :
Db    188 SVSSTSSSTFSSAAPTSTSSSYLSSSSVVSSSSSPSSSSSSTLTSSSLSTSSIPSTSSSS 247

Qy    316 SS 317
          ||
Db    248 SS 249


RESULT   6
I64138
adhesin homolog HI1732 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Oct-1997
C;Accession: I64138
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: I64138
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-298 <TIGR>
A;Cross-references: GB:U32846; GB:L42023; NID:g1574588; PID:g1574589; TIGR:HI1732


  Query Match             5.9%;  Score 99;  DB 2;  Length 298;
  Best Local Similarity   22.8%;  Pred. No. 3;
  Matches   66;  Conservative   39;  Mismatches  100;  Indels   84;  Gaps   15;

Qy     11 FVTV-LTFNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTAQ 69
          || |  | :   ||   :: |  : |   |:     || :   :|:  |   |:| | |:
Db     56 FVKVQSTEDDIEDSAATKDDNKNQALKAGDTLTLKAGKNLKAKLDQ--GGKSVTFAL-AK 112

Qy     70 FYEIKPTKENEQY-IGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTKE 128
            ::|  | ::   ||    |    :|             ||   |  |        | |
Db    113 DLDVKTAKVSDTLTIGGNTPAAGGATP----------KVSITSTAD----GLKLAKG 155

Qy    129 VKGQT----NKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKM 184
            | |    | ||| :|||                |:| :|  :||  | |    ||  :
Db    156 TNGDTAVHLNGLASTLPDVT--------------TNTGASTSVTFS-PSDIEKTRAATI 199

Qy    185 RQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVT-EKRIIDPSLFS 243
          : :| | |:: :       | : |      : :| ::  | |   :| :|  :|  | :
Db    200 KDVL-NAGWNIK------GAKVAGGN-----TENVDLVAGYDNVEFITGDKNTLDVVLTA 247

Qy    244 --SGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNN 290
            :|  |:  :                          :|  | : |||
Db    248 KENGKTTEVKF--------------------------TPKTSVIKDNN 269


RESULT   7
JQ1792
SalF17R protein - vaccinia virus
N;Alternate names: SalF21R 64.7K protein
C;Species: vaccinia virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 22-Oct-2001
C;Accession: JQ1792; C38550
R;Smith, G.L.; Chan, Y.S.; Howard, S.T.
J. Gen. Virol. 72, 1349-1376, 1991
A;Title: Nucleotide sequence of 42kbp of vaccinia virus strain WR from near the right
A;Reference number: JQ1767; MUID:91259063
A;Accession: JQ1792
A;Molecule type: DNA
A;Residues: 1-564 <SMI>
A;Cross-references: DDBJ:D11079; NID:g222717; PIDN:BAA01828.1; PID:g222743
A;Experimental source: strain WR
R;Howard, S.T.; Chan, Y.S.; Smith, G.L.
Virology 180, 633-647, 1991
A;Title: Vaccinia virus homologues of the shope fibroma virus inverted terminal repea
A;Reference number: A38550; MUID:91111982
A;Accession: C38550
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-564 <HOW>
A;Cross-references: GB:M58052
A;Experimental source: strain WR
C;Superfamily: A55R protein; POZ domain homology
F;7-110/Domain: POZ domain homology <POZ>


  Query Match             5.9%;  Score 99;  DB 2;  Length 564;
  Best Local Similarity   22.1%;  Pred. No. 7.2;
```

```
  Matches   73;  Conservative   50;  Mismatches  109;  Indels   98;  Gaps   18;

Qy      11 FVTVLTFNSCADSNGNQEIN--GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFML-T 67
           : ::|  |:  |||   |:|    :  ||||  | |:   :|: :  :|    | ::| |
Db      45 YFSILFSNNFIDSN-EYEVNLSHLDYQSVND--LIDYIYGIPLSLTNDN----VKYILST 97

Qy      68 AQFYEI-KPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTIL- 125
           | | :|     | | || :    ||    :|:    :|:  |:||||      | |||
Db      98 ADFLQIGSAITECENYI-LKNLCSKNCIDFYIYADKYNNK--KIESAS------FNTILQ 148

Qy     126 --TKEVKGQTNKLASVIPDVATLNSLFNQIKN----------------QSCGTS------ 161
             : :  :  |  :    :  |:     |||                  |||
Db     149 NILRLINDENFKYLTEESMIKILSDDMLNIKNEDFAPLILIKWLESTQQSCTVELLRCLR 208

Qy     162 -TASSP----------------CITFRYPV---DGCYARAHKMR--QILMNNGYDCEKQ 198
            :  ||                |||||   :   |  : | | :    | ::| :|
Db     209 ISLLSPQVIKSLYSHQLVSSIYECITFLNNIAFLDESFPRYHSIELISIGISNSHDKISI 268

Qy     199 FVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVN 258
           | | | :|     :  | :  |:            ::|  ::  |    : :|:: |
Db     269 NCY-NHKKNTWEMISSRRYRCSFAVA----------VLDNIIYMMGGYDQSPYRSSKV- 315

Qy     259 TSCGSASVSSYANTAGNVYYRSPSNSYLYD 288
                       : | : :||::||
Db     316 ----------------IAYNTCTNSWIYD 328


RESULT   8
T50765
adhesion of calyx edges protein ACE [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T50765
R;Araki, T.; Nakatani-Goto, M.
submitted to the EMBL Data Library, May 1999
A;Description: ADHESION OF CALYX EDGES (ACE), cDNA.
A;Reference number: Z25224
A;Accession: T50765
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-594 <ARA>
A;Cross-references: EMBL:AB027458; PIDN:BAA77837.1
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: ACE
A;Map position: 1


  Query Match             5.9%;  Score 99;  DB 2;  Length 594;
  Best Local Similarity   21.0%;  Pred. No. 7.7;
  Matches   49;  Conservative   28;  Mismatches   74;  Indels   82;  Gaps   11;

Qy      81 QYIGMLRQAVK-------NESP----VHIFLKPNSNEIGKVESASPEDVRYFKTILTKEV 129
           | :|: :  |          :||      |  :  | ||:             | ||  |:
Db     353 QTVGITKMGVYVEASTGFGQSPESIHTHYGIMSNKNEL------------FSTIPAKQR 399

Qy     130 KGQTNKLASVIPDVATLNSLFN----------------QIKNQSCGTSTASSPCITFRY 172
           : :  : | :  :   |:  ||                : |    |:   :| :|| |
Db     400 RPEATQ-AYITRNKYQLHEAFNGSFILEKLAYPISRGHLSLVN----TNVDDNPSVTFNY 454

Qy     173 ---PVD--GCYARAHKMRQILMNNGY----DCEKQFVYGNLKAST--------------- 208
             |||   |       : ::: :| :      |:|| |:  |  |
Db     455 FKHPVDLQRCVEAIRLVSKVVTSNRFLNYTQCDKQNVHKMLSLSVKANINLRPKQLNDTK 514

Qy     209 -------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFSSGPVTD 249
                  |    | ||   ||        :   ||   |:|| | |   | |:
Db     515 SMAQFCKDTVVTIWHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFDESPGTN 567


RESULT   9
C83912
hypothetical protein BH2099 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: C83912
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: C83912
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-419 <STO>
A;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05818.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH2099


  Query Match             5.9%;  Score 98.5;  DB 2;  Length 419;
  Best Local Similarity   23.0%;  Pred. No. 5.2;
  Matches   65;  Conservative   38;  Mismatches   99;  Indels   81;  Gaps   12;

Qy      12 VTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTAQFY 71
           | || |       |:||:|    |:|              |  |:| | |       ||
Db      30 VIVLNFAFAGLFGGDQELNLDLTLAV------------VNEDDETGTI-------AQLQ 69

Qy      72 E--IKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTI----- 124
           |  ||     |  :  | : | :  |: :            |  | ||:: : |:
Db      70 EKLIKEASFKETEVARLVEQVGDIQPIPLLF----------EYLSSEDLQDWLTVYELTE 119

Qy     125 ---LTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARA 181
             : |  :|  : :  ||||  |:|||:  :      |  ||||      | |:   : :
Db     120 AEAVAKVKEGAADGIL-VIPDGFTVNSLYAAL----VGEPTASS----LIYKVEMETSNS 170

Qy     182 HKMRQIL------MNNGYDCEKQFVYGNLKASTG------------------TCCVAWS 216
             : ||:      ||| :  ::        |  |                  :    ::
Db     171 ATLYQIIDGFLDHMNNHFALQQMGEVQEANAPEGGLEKIGVGEAFTLTQYFTISMSALFA 230

Qy     217 YHVAILVSYKNASGVTEK---RII----DPSLFSSGPVTDTAW 252
             :|  |: |    : :|   ||:     | ||  | :  ||:
Db     231 LFLAATVATKTGEEIRQKVFHRILLTDRKPILFLVGKIVSTAF 273


RESULT  10
C69146
hypothetical protein MTH359 - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: C69146
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: C69146
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-534 <MTH>
A;Cross-references: GB:AE000821; GB:AE000666; NID:g2621414; PIDN:AAB84865.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH359


  Query Match             5.8%;  Score 97;  DB 2;  Length 534;
  Best Local Similarity   20.9%;  Pred. No. 9.6;
  Matches   57;  Conservative   34;  Mismatches  108;  Indels   74;  Gaps    9;

Qy      44 DFGKTVPVGI----------DEENGMIKVSFMLTAQ----FYEIKPTKENEQYIGMLRQ 88
           : ||  ||             |    : :: ::  |     | |      |  | |  :
Db     277 NLGKNTPVAPVTAGEAANPAGDARGTLTRLEYLKVASGVQGFMEKNHRAPN--YAGSSKG 334

Qy      89 AVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNS 148
             :  || |:         | :| |    : |                    :|:  |:
```

```
Db    335 RISYESLVY--------SISRVLSFYSSNKR--------------------LPNTVTVTR 366

Qy    149 LFNQIKNQS-----CGTSTASSPCITFRYPVD----------------GCYARAHK---- 183
          | : :||:       | |||     :   |||                 ::||
Db    367 LASSLKNRPENDPYRGESTARYLASSASCPVDSPDIRSLASEITRGLTSTFSRAEAVFGW 426

Qy    184 MRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFS 243
          :|  :  : |   |    | ||  || |      |  :||:   |:|:  : :     |:
Db    427 VRDNINYSFYYNTKYGAVGTLKNRTGNCV----DHTHLLVALARAAGIPARYVHGTCNFT 482

Qy    244 SGPVTDTAWRNACVNTSCGSASVSSYANTAGNV 276
          || |    |    |  :  :|  :|  |: | |
Db    483 SGNVYGHVWAQLLVGDTWYAADATSSRNSLGVV 515


RESULT  11
B90026
hypothetical protein modA [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: B90026
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: B90026
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-260 <KUR>
A;Cross-references: GB:BA000018; PID:g13702079; PIDN:BAB43371.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: modA
C;Superfamily: molybdate-binding periplasmic protein


  Query Match             5.7%;  Score 96.5;  DB 2;  Length 260;
  Best Local Similarity   22.5%;  Pred. No. 3.9;
  Matches   42;  Conservative   39;  Mismatches   59;  Indels   47;  Gaps    9;

Qy      1 MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDF--GKTVPVGIDEENG 58
          ||     :||   ||    |::|| |           |:||  |   ||   : :
Db      3 MKRFIAIVMALFLVLA--GCSNSNDN----------NESKKDDADNGKKQEIQVAAAAS 49

Qy     59 MIKVSFMLTAQF---YEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIG--KVESA 113
          :  |:  | ::|   ::    | |    | ||: ::: :|| :|:  |: ::   | ::
Db     50 LTDVTKKLASEFKKEHKNADIKFNYGGSGALRKQIESGAPVDVFMSANTKDVDALKDKNK 109

Qy    114 SPEDVRYFKTIL----------TKEVKGQTNKLASVIPDVATL----------NSL 149
          : :  :| |  |           |::|   :|||  : :| |:           |:|
Db    110 AHDTYKYAKNSLVLIGDKDSNYTSVKDLK-DNDKLA--LGEVKTVPAGKYAKQYLDNNNL 166

Qy    150 FNQIKNQ 156
          | :::::
Db    167 FKEVESK 173


RESULT  12
H70103
hypothetical protein BB0032 - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: H70103
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: H70103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-490 <KLE>
A;Cross-references: GB:AE001117; GB:AE000783; NID:g2687907; PIDN:AAC66427.1; PID:g268
A;Experimental source: strain B31


  Query Match             5.7%;  Score 96.5;  DB 2;  Length 490;
  Best Local Similarity   26.0%;  Pred. No. 9.3;
  Matches   50;  Conservative   31;  Mismatches   70;  Indels   41;  Gaps   10;

Qy      5 FLSMMAFVTVLTFNSCADSNGNQEINGKEKL----SVNDSKLKDFGKTVPVGIDEE-NGM 59
          ||     ||   :      |    :| || |     |::: ||:| :   :|::|   |:
Db    309 FLIQEQNVTKYEYQDFLKENPKWALNNKENLIKEQLVDENYLKNFNQ---IGLNEAITGI 365

Qy     60 IKVSFMLTAQFYEIK-PT---------KENEQYIGMLRQAVKNESPVHIFLKPNSNEIGK 109
            | :  | :|  | ||          :| | |     |   |::|::|      ||| |
Db    366 SYFSAIEYANWYSKKLPTGFKARLPISQEWELY-----QKEPNKNPLNI------NEISK 414

Qy    110 -------VESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTS- 161
                ::::|  ::  ||    |    : :   |:| :: | :   | : | |   |
Db    415 KVGFWNLMQNSSFNEIAIFKN--EKNFYSENSNFYSLITEIRTYSQQNNNLLNASTKASF 472

Qy    162 --TASSPCITFR 171
              ||| | ||
Db    473 LKNWSSPNIGFR 484


RESULT  13
S49802
probable membrane protein YML059c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM9958.03c
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 05-Nov-1999
C;Accession: S49802
R;Devlin, K.; Churcher, C.
submitted to the EMBL Data Library, November 1994
A;Reference number: S49800
A;Accession: S49802
A;Molecule type: DNA
A;Residues: 1-1679 <DEV>
A;Cross-references: EMBL:Z46729; NID:g577134; PID:g577137; GSPDB:GN00013; MIPS:YML059
C;Genetics:
A;Gene: MIPS:YML059c
A;Map position: 13L
C;Keywords: transmembrane protein
F;103-119/Domain: transmembrane #status predicted <TM1>
F;235-251/Domain: transmembrane #status predicted <TM2>


  Query Match             5.7%;  Score 95.5;  DB 2;  Length 1679;
  Best Local Similarity   22.8%;  Pred. No. 60;
  Matches   77;  Conservative   42;  Mismatches  116;  Indels  103;  Gaps   16;

Qy     17 FNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTAQFYEIKPT 76
          | :   |:|: ::  | | |   ||||   |  |      :|:|:           :|
Db    492 FKTVDKSSGSADLEPKPKNSNASSKLKKPPKAKP-----SDGIIQ----------SLK-- 534

Qy     77 KENEQYIGMLRQAVKNESPVHIFLKPNSNE--------IGKVESASPEDVRYFKTILTKE 128
                :  |  | |   : |||            :|  :  :| |
Db    535 ---------IANANANTSSNSLSLKPEFTHHPSSRHVVLGSRDQFNPGD---------- 574

Qy    129 VKGQTNKLASVIP-----DVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHK 183
              | | :|     |: : | : |  :|:| | :|::|                 ||
Db    575 -------LLSNVPLSRTMDILSPNPIHNNNRNKSNGINTSTS-------------NQHK 613

Qy    184 -MRQILMNNGYDCEKQF-----VYGNLKASTGTCCV-AWSYHVAILVSYKNASGVTEKR- 235
           :   ||     |:|        | : ||    :   | |  |  |  :  |    |
Db    614 RSSRSSSNNASVHSKKFSSLSPELRNAQLSTSPLSLDNTSVHDHIHPSPVHLKGRVSPRP 673

Qy    236 -IIDPSLFSSG--PVTDTAWRNAC---------VNTSCGSASVSSYANTA-------GNV 276
           ::  : ||:      |:| | |           || |  | | || || :         :
Db    674 NLLPTTSFSAAQEETEDSALRMALVEAMLTYLGVNKSNMSVSSSSIANMSSLNSPQLNEM 733
```

```
Qy    277 YYRSPSNSYLYDNNLINTNCVLTKFSLLSG-CSPSPAP 313
Db    734 YSRRPSNA----SFLMSPHCTPSDISVASSFASPQTQP 767


RESULT  14
C42523
A55R protein - vaccinia virus (strain Copenhagen)
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 22-Oct-2001
C;Accession: C42523
R;Johnson, G.P.
submitted to GenBank, June 1990
A;Reference number: A33172
A;Accession: C42523
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-564 <JOH>
C;Superfamily: A55R protein; POZ domain homology
F;7-110/Domain: POZ domain homology <POZ>


  Query Match             5.7%;  Score 95;  DB 2;  Length 564;
  Best Local Similarity   22.1%;  Pred. No. 15;
  Matches   73;  Conservative   49;  Mismatches  110;  Indels   98;  Gaps   18;

Qy     11 FVTVLTFNSCADSNGNQEIN--GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFML-T 67
Db     45 YFSILFSNNFIDSN-EYEVNLSHLDYQSVND--LIDYIYGIPLSLTNDN----VKYILST 97

Qy     68 AQFYEI-KPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTIL- 125
Db     98 ADFLQIGSAITECENYI-LKNLCSKNCIDFYIYADKYNNK--KIESAS------FNTILQ 148

Qy    126 --TKEVKGQTNKLASVIPDVATLNSLFNQIKN----------------QSCGTS------ 161
Db    149 NILRLINDENFKYLTEESMIKILSDDMLNIKNEDFAPLILIKWLESTQQSCTVELLKCLR 208

Qy    162 -TASSP-----------------CITFRYPV---DGCYARAHKMR--QILMNNGYDCEKQ 198
Db    209 ISLLSPQVIKSLYSHRLVSSIYECITFLNNIAFLDESFPRYHSIELISIGISNSRDKISI 268

Qy    199 FVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVN 258
Db    269 NCY-NHKKNTWEMISSRRYRCSFAVA----------VLDNIIYMMGGYDQSPYRSSKV- 315

Qy    259 TSCGSASVSSYANTAGNVYYRSPSNSYLYD 288
Db    316 ----------------IAYNTCTNSWIYD 328


RESULT  15
A37023
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenz
N;Alternate names: immunoglobulin A1 proteinase type 1
C;Species: Haemophilus influenzae
C;Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000
C;Accession: A37023
R;Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thogersen, H.C.; Kilian, M.
Infect. Immun. 57, 3097-3105, 1989
A;Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of Haemophi
A;Reference number: A37023; MUID:89379374
A;Accession: A37023
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1541 <POU>
A;Cross-references: GB:X64357; NID:g43560; PIDN:CAA45708.1; PID:g43561
A;Experimental source: serotype b
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


  Query Match             5.7%;  Score 95;  DB 2;  Length 1541;
  Best Local Similarity   23.5%;  Pred. No. 59;
  Matches   65;  Conservative   39;  Mismatches  101;  Indels   72;  Gaps   17;

Qy     17 FNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTAQFYEIKPT 76
Db     37 FRDFAENKGKFSV-GATNVLVKDKNNKDLGTALP------NGIPMIDFSVVDVDKRI-AT 88

Qy     77 KENEQYIGMLRQAVKNESPVHI-FLKPNSNEIGKVES---ASPEDVRYF----------- 121
Db     89 LINPQYVVGVKHVSNGVSELHFGNLNGNMNN-GNAKAHRDVSSEENRYFSVEKNEYPTKL 147

Qy    122 --KTILTKEVKGQTNK---------LASVIPDVATLNSLFNQIKNQSCGTSTASSPCITF 170
Db    148 NGKTVTTED---QTQKRREDYYMPRLDKFVTEVAPIEA-----------STASSDAGTY 192

Qy    171 ----RYPVDGCYARAHKMRQILMNNGYD----CEKQFVYGNLKASTGTCCVAWSYHVA-- 220
Db    193 NDQNKYP---AFVRLGSGSQFIYKKGDNYSLILNNHEVGGNNLKLVGD---AYTYGIAGT 246

Qy    221 -ILVSYKN----ASGVTEKRIIDP-SLFSSGPVTDTA 251
Db    247 PYKVNHENNGLIGFGNSKEEHSDPKGILSQDPLTNYA 283


Search completed: June 27, 2002, 21:49:27
Job time: 5215 sec
```



GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:59:19 ; Search time 39.56 Seconds
(without alignments)
313.202 Million cell updates/sec

Title: US-09-727-769A-8
Perfect score: 1679
Sequence: 1 MKNLFLSMMAFVTVLTFNSC.....FSLLSGCSPSPAPDVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	99.5	5.9	527	1	VG12_BPT4	P10930 bacteriopha
2	99	5.9	564	1	VA55_VACCV	P24768 vaccinia vi
3	96.5	5.7	490	1	Y032_BORBU	O51063 borrelia bu
4	95.5	5.7	1679	1	YMF9_YEAST	Q04958 saccharomyc
5	95	5.7	564	1	VA55_VACCC	P21073 vaccinia vi
6	95	5.7	1541	1	IGA1_HAEIN	P42782 haemophilus
7	94.5	5.6	622	1	YF54_METJA	Q58949 methanococc
8	94.5	5.6	850	1	D7_DICDI	P54682 dictyosteli
9	93	5.5	1183	1	CNA_STAAU	Q53654 staphylococ
10	92	5.5	344	1	CYS5_DICDI	P54640 dictyosteli
11	91.5	5.4	851	1	NUD1_YEAST	P32336 saccharomyc
12	91.5	5.4	3130	1	DPOZ_HUMAN	O60673 homo sapien
13	91	5.4	761	1	NCA2_HUMAN	P13592 homo sapien
14	91	5.4	848	1	NCA1_HUMAN	P13591 homo sapien
15	89	5.3	1895	1	YLK3_CAEEL	P41951 caenorhabdi
16	88.5	5.3	1251	1	YQU3_CAEEL	Q09550 caenorhabdi
17	88	5.2	386	1	RMAR_HANWI	P48849 hansenula w
18	88	5.2	458	1	ME31_DROME	P23128 drosophila
19	87.5	5.2	669	1	MTR1_MOUSE	Q9z2c4 mus musculu
20	87.5	5.2	928	1	PMP9_CHLPN	Q9z398 chlamydia p
21	87	5.2	817	1	ENPL_CATRO	P35016 catharanthu
22	86.5	5.2	351	1	MREC_HAEIN	P44475 haemophilus
23	86.5	5.2	657	1	RES2_SCHPO	P41412 schizosacch
24	86.5	5.2	821	1	YQIG_ECOLI	P76655 escherichia
25	86	5.1	383	1	FMA_PORGI	P13793 porphyromon
26	86	5.1	888	1	LAGC_DICDI	P42523 dictyosteli
27	85.5	5.1	638	1	GHR_RABIT	P19941 oryctolagus
28	85.5	5.1	1082	1	SP23_YEAST	P35210 saccharomyc
29	85.5	5.1	2715	1	G156_PARPR	P13837 paramecium
30	84.5	5.0	443	1	TOLB_RICPR	Q9zdm5 rickettsia
31	84.5	5.0	666	1	YEA7_YEAST	P40002 saccharomyc
32	84.5	5.0	690	1	TRFE_ORYLA	P79819 oryzias lat
33	84.5	5.0	1401	1	LATA_LATMA	P23631 latrodectus
34	84.5	5.0	1620	1	DNM1_MOUSE	P13864 mus musculu
35	84.5	5.0	2704	1	G168_PARPR	P17053 paramecium
36	84	5.0	533	1	CAG2_MOUSE	Q09200 mus musculu
37	84	5.0	725	1	NCA2_MOUSE	P13594 mus musculu
38	84	5.0	1115	1	NCA1_MOUSE	P13595 mus musculu
39	84	5.0	1694	1	IGA0_HAEIN	P44969 haemophilus
40	83.5	5.0	405	1	ASSY_THEAC	Q9hkf1 thermoplasm
41	83.5	5.0	537	1	VGLF_PVM	P35949 pneumonia v
42	83.5	5.0	594	1	CIK1_YEAST	Q01649 saccharomyc
43	83.5	5.0	775	1	UL87_HSV7J	P52363 human herpe
44	83	4.9	431	1	YG84_METTH	O27719 methanobact
45	83	4.9	435	1	FUZ7_USTMA	Q99078 ustilago ma

ALIGNMENTS

RESULT 1
VG12_BPT4
ID VG12_BPT4 STANDARD; PRT; 527 AA.
AC P10930;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Short tail fiber protein (Protein Gp12) (p12).
GN 12.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D;
RX MEDLINE=88189824; PubMed=3357780;
RA Selivanov N.A., Prilipov A.G., Mesyanzhinov V.V.;
RT "Nucleotide and deduced amino acid sequence of bacteriophage T4 gene 12.";
RL Nucleic Acids Res. 16:2334-2334(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 1-57 FROM N.A.
RC STRAIN=D;
RX MEDLINE=89356257; PubMed=2548819;
RA Barrett B.K., Berget P.B.;
RT "Using transposon Tn5 insertions to sequence bacteriophage T4 gene 11.";
RL DNA 8:287-295(1989).
RN [4]
RP SEQUENCE OF 1-4 FROM N.A.
RC STRAIN=D;
RX MEDLINE=89263746; PubMed=2726468;
RA Prilipov A.G., Selivanov N.A., Efimov V.P., Marusich E.I.,
RA Mesyanzhinov V.V.;
RT "Nucleotide sequences of bacteriophage T4 genes 9, 10 and 11.";
RL Nucleic Acids Res. 17:3303-3303(1989).
CC -!- FUNCTION: STRUCTURAL COMPONENT OF THE SHORT TAIL FIBER (BASEPLATE).
CC ---

CC ---
DR EMBL; X06792; CAA29951.1; -.
DR EMBL; AF158101; AAD42417.1; -.

```
DR   EMBL; M26253; AAA32495.1; -.
DR   EMBL; X14192; CAA32398.1; -.
DR   PIR; S01889; GIBPT4.
DR   PIR; C32479; C32479.
KW   Fiber protein.
FT   CONFLICT     33     33       H -> Q (IN REF. 3).
SQ   SEQUENCE   527 AA;  56214 MW;  808FF05DC3DABB7D CRC64;


  Query Match             5.9%;  Score 99.5;  DB 1;  Length 527;
  Best Local Similarity   20.8%;  Pred. No. 1.6;
  Matches   77;  Conservative   45;  Mismatches  134;  Indels  115;  Gaps   15;

Qy      24 NGNQEINGKEKLSVNDSKLKDFGK---------TVP--------VGIDEENGMIKVSFML 66
           |  : : |  : | ||  :              ||         |  :  ||:||:| :
Db      89 NATETVYGLTRYSTNDEAIAGVNNESSITPAKFTVALNNAFETRVSTESSNGVIKISSLP 148

Qy      67 TA-----QFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVR-- 119
            |          : | |  :  | :: |   :|:          :: | |: |:   ||
Db     149 QALAGADDTTAMTPLKTQQLAIKLIAQIAPSETTA------TESDQGVVQLATVAQVRQG 202

Qy     120 ------------YFKTILTKEVKG---------QTNKLASVIPDVATLN------SLFNQ 152
                       :  :  |:| ||          :  |||    ||||      |:
Db     203 TLREGYAISPYTFMNSSSTEEYKGVIKLGTQSEVNSNNASVAVTGATLNGRGSTTSMRGV 262

Qy     153 IKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKAS----- 207
           :|     |:|| |     :   |   | |     |    |     | :|| |:
Db     263 VKL----TTTAGS-----QSGGDASSALAWNADVIQQRGG-----QIIYGTLRIEDTFTI 308

Qy     208 -------TGTCCVAWSYHVAILVSYKNASGVTEKRIIDPS-------LFSSGPVTDTAWR 253
                  |||         | :   |   : :  :  || :        ::::  :   |||
Db     309 ANGGANITGT--------VRMTGGYIQGNRIVTQNEIDRTIPVGAIMMWAADSLPSDAWR 360

Qy     254 NACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYD----------------NNLINTNCV 297
             |   :  ::    ||:  |  |  :|||  | |                 |  :| |
Db     361 -FCHGGTVSASDCPLYASRIGTRYGGNPSNPGLPDMRGLFVRGSGRGSHLTNPNVNGNDQ 419

Qy     298 LTKFSLLSGCS 308
             |  ~|  ||:
Db     420 FGKPRLGVGCT 430


RESULT   2
VA55_VACCV
ID   VA55_VACCV     STANDARD;      PRT;   564 AA.
AC   P24768;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A55.
GN   A55R OR SALF17R OR SALF21R.
OS   Vaccinia virus (strain WR).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91259063; PubMed=2045793;
RA   Smith G.L., Chan Y.S., Howard S.T.;
RT   "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT   the right inverted terminal repeat.";
RL   J. Gen. Virol. 72:1349-1376(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91111982; PubMed=1846491;
RA   Howard S.T., Chan Y.S., Smith G.L.;
RT   "Vaccinia virus homologues of the Shope fibroma virus inverted
RT   terminal repeat proteins and a discontinuous ORF related to the tumor
RT   necrosis factor receptor family.";
RL   Virology 180:633-647(1991).
CC   -!- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES KELCH FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D11079; BAA01828.1; -.
DR   EMBL; M58054; AAA48340.1; -.
DR   PIR; C38550; C38550.
DR   PIR; JQ1792; JQ1792.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR001798; Kelch.
DR   Pfam; PF00651; BTB; 1.
DR   Pfam; PF01344; Kelch; 6.
DR   PRINTS; PR00501; KELCHREPEAT.
DR   SMART; SM00225; BTB; 1.
DR   PROSITE; PS50097; BTB; 1.
KW   Early protein; Repeat.
FT   DOMAIN       21     88       BTB.
FT   REPEAT      252    297       KELCH 1.
FT   REPEAT      298    346       KELCH 2.
FT   REPEAT      347    395       KELCH 3.
FT   REPEAT      397    441       KELCH 4.
FT   REPEAT      442    492       KELCH 5.
FT   REPEAT      494    539       KELCH 6.
SQ   SEQUENCE   564 AA;  64690 MW;  FDAC2001F6EF7507 CRC64;


  Query Match             5.9%;  Score 99;  DB 1;  Length 564;
  Best Local Similarity   22.1%;  Pred. No. 1.9;
  Matches   73;  Conservative   50;  Mismatches  109;  Indels   98;  Gaps   18;

Qy      11 FVTVLTFNSCADSNGNQEIN--GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFML-T 67
           : ::|  |:  |||   |:|    :  |||| | |:   :|: :  :|    | ::| |
Db      45 YFSILFSNNFIDSN-EYEVNLSHLDYQSVND--LIDYIYGIPLSLTNDN----VKYILST 97

Qy      68 AQFYEI-KPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTIL- 125
           | | :|     | | || :     ||    :|:    :|:  |:|||      | |||
Db      98 ADFLQIGSAITECENYI-LKNLCSKNCIDFYIYADKYNNK--KIESAS------FNTILQ 148

Qy     126 --TKEVKGQTNKLASVIPDVATLNSLFNQIKN---------------QSCGTS------ 161
             : :  :  |  :    :  |:    |||                 |||
Db     149 NILRLINDENFKYLTEESMIKILSDDMLNIKNEDFAPLILIKWLESTQQSCTVELLRCLR 208

Qy     162 -TASSP----------------CITFRYPV---DGCYARAHKMR--QILMNNGYDCEKQ 198
            :  ||                |||   :   |  : | | :    | ::| :|
Db     209 ISLLSPQVIKSLYSHQLVSSIYECITFLNNIAFLDESFPRYHSIELISIGISNSHDKISI 268

Qy     199 FVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVN 258
           | | | :|     :  |  :  |:          ::|  ::  |    : :|:: |
Db     269 NCY-NHKKNTWEMISSRRYRCSFAVA----------VLDNIIYMMGGYDQSPYRSSKV- 315

Qy     259 TSCGSASVSSYANTAGNVYYRSPSNSYLYD 288
                       : | : :||::||
Db     316 ----------------IAYNTCTNSWIYD 328


RESULT   3
Y032_BORBU
ID   Y032_BORBU     STANDARD;      PRT;   490 AA.
AC   O51063;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein BB0032.
GN   BB0032.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001117; AAC66427.1; -.
DR   TIGR; BB0032; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     27     47       POTENTIAL.
SQ   SEQUENCE   490 AA;  57590 MW;  0A4FD696A27E5548 CRC64;


  Query Match             5.7%;  Score 96.5;  DB 1;  Length 490;
  Best Local Similarity   26.0%;  Pred. No. 2.5;
  Matches   50;  Conservative   31;  Mismatches   70;  Indels   41;  Gaps   10;

Qy       5 FLSMMAFVTVLTFNSCADSNGNQEINGKEKL----SVNDSKLKDFGKTVPVGIDEE-NGM 59
           ||     ||   :       |    :| || |     |::: ||:| :   :|::|   |:
Db     309 FLIQEQNVTKYEYQDFLKENPKWALNNKENLIKEQLVDENYLKNFNQ---IGLNEAITGI 365

Qy      60 IKVSFMLTAQFYEIK-PT---------KENEQYIGMLRQAVKNESPVHIFLKPNSNEIGK 109
             | :  | :|  | ||         :| | |      |   |::|::|      ||| |
Db     366 SYFSAIEYANWYSKKLPTGFKARLPISQEWELY-----QKEPNKNPLNI------NEISK 414

Qy     110 -------VESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTS- 161
                 ::::|  ::  ||    |    : :   |:| :: | :   | : | |   |
Db     415 KVGFWNLMQNSSFNEIAIFKN--EKNFYSENSNFYSLITEIRTYSQQNNNLLNASTKASF 472

Qy     162 --TASSPCITFR 171
               ||| | ||
Db     473 LKNWSSPNIGFR 484


RESULT   4
YMF9_YEAST
ID   YMF9_YEAST     STANDARD;      PRT;  1679 AA.
AC   Q04958;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 187.1 kDa protein in OGG1-CNA2 intergenic region.
GN   YML059C OR YM9958.03C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Devlin K., Churcher C., Barrell B.G., Rajandream M.A.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46729; CAA86716.1; -.
DR   SGD; S0004524; YML059C.
DR   InterPro; IPR001423; UPF0028.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 2.
DR   Pfam; PF01173; UPF0028; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 2.
DR   PROSITE; PS01237; UPF0028; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     50     70       POTENTIAL.
FT   TRANSMEM    104    124       POTENTIAL.
SQ   SEQUENCE   1679 AA;  187132 MW;  550FFCD4ACAF8E25 CRC64;


  Query Match             5.7%;  Score 95.5;  DB 1;  Length 1679;
  Best Local Similarity   22.8%;  Pred. No. 14;
  Matches   77;  Conservative   42;  Mismatches  116;  Indels  103;  Gaps   16;

Qy      17 FNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTAQFYEIKPT 76
           | :   |:|: ::  | | |   ||||   |  |      :|:|:           :|
Db     492 FKTVDKSSGSADLEPKPKNSNASSKLKKPPKAKP-----SDGIIQ----------SLK-- 534

Qy      77 KENEQYIGMLRQAVKNESPVHIFLKPNSNE--------IGKVESASPEDVRYFKTILTKE 128
                 :  |  | |   : |||             :|  :  :| |
Db     535 ---------IANANANTSSNSLSLKPEFTHHPSSRHVVLGSRDQFNPGD---------- 574

Qy     129 VKGQTNKLASVIP-----DVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHK 183
                 | | :|     |: : | : |  :|:| | :|::|                ||
Db     575 -------LLSNVPLSRTMDILSPNPIHNNNRNKSNGINTSTS-------------NQHK 613

Qy     184 -MRQILMNNGYDCEKQF-----VYGNLKASTGTCCV-AWSYHVAILVSYKNASGVTEKR- 235
             :  ||     |:|        | : ||    :   | |  |  |  :  |    |
Db     614 RSSRSSSNNASVHSKKFSSLSPELRNAQLSTSPLSLDNTSVHDHIHPSPVHLKGRVSPRP 673

Qy     236 -IIDPSLFSSG--PVTDTAWRNAC---------VNTSCGSASVSSYANTA-------GNV 276
            ::  : ||:      |:| | |             || |  | | || || :       :
Db     674 NLLPTTSFSAAQEETEDSALRMALVEAMLTYLGVNKSNMSVSSSSIANMSSLNSPQLNEM 733

Qy     277 YYRSPSNSYLYDNNLINTNCVLTKFSLLSG-CSPSPAP 313
           | | |||:    : |:: :|  :  |: |   ||   |
Db     734 YSRRPSNA----SFLMSPHCTPSDISVASSFASPQTQP 767


RESULT   5
VA55_VACCC
ID   VA55_VACCC     STANDARD;      PRT;   564 AA.
AC   P21073;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A55.
GN   A55R.
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
```

```
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
CC   -!- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES KELCH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35027; AAA48190.1; -.
DR   PIR; C42523; C42523.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR001798; Kelch.
DR   Pfam; PF00651; BTB; 1.
DR   Pfam; PF01344; Kelch; 6.
DR   PRINTS; PR00501; KELCHREPEAT.
DR   SMART; SM00225; BTB; 1.
DR   PROSITE; PS50097; BTB; 1.
KW   Early protein; Repeat.
FT   DOMAIN       21     88       BTB.
FT   REPEAT      252    297       KELCH 1.
FT   REPEAT      298    346       KELCH 2.
FT   REPEAT      347    395       KELCH 3.
FT   REPEAT      397    441       KELCH 4.
FT   REPEAT      442    492       KELCH 5.
FT   REPEAT      494    539       KELCH 6.
SQ   SEQUENCE   564 AA;  64718 MW;  DBE8ECB532DA9F71 CRC64;


  Query Match             5.7%;  Score 95;  DB 1;  Length 564;
  Best Local Similarity   22.1%;  Pred. No. 4;
  Matches   73;  Conservative   49;  Mismatches  110;  Indels   98;  Gaps   18;

Qy      11 FVTVLTFNSCADSNGNQEIN--GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFML-T 67
           : ::|  |:  |||   |:|    :  ||||  | |:   :|: :  :|    | ::| |
Db      45 YFSILFSNNFIDSN-EYEVNLSHLDYQSVND--LIDYIYGIPLSLTNDN----VKYILST 97

Qy      68 AQFYEI-KPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTIL- 125
           | | :|     | | || :     ||    :|:    :|:  |:||||      | |||
Db      98 ADFLQIGSAITECENYI-LKNLCSKNCIDFYIYADKYNNK--KIESAS------FNTILQ 148

Qy     126 --TKEVKGQTNKLASVIPDVATLNSLFNQIKN----------------QSCGTS------ 161
             : :  :  |  :    :  |:     |||                 |||
Db     149 NILRLINDENFKYLTEESMIKILSDDMLNIKNEDFAPLILIKWLESTQQSCTVELLKCLR 208

Qy     162 -TASSP----------------CITFRYPV---DGCYARAHKMR--QILMNNGYDCEKQ 198
            :  ||                |||||   :   |  : | | :    | ::|  |
Db     209 ISLLSPQVIKSLYSHRLVSSIYECITFLNNIAFLDESFPRYHSIELISIGISNSRDKISI 268

Qy     199 FVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVN 258
           | | | :|     :  |  :  |:           ::|  ::  |    : :|:: |
Db     269 NCY-NHKKNTWEMISSRRYRCSFAVA----------VLDNIIYMMGGYDQSPYRSSKV- 315

Qy     259 TSCGSASVSSYANTAGNVYYRSPSNSYLYD 288
                     : | : :||::||
Db     316 ----------------IAYNTCTNSWIYD 328


RESULT   6
IGA1_HAEIN
ID   IGA1_HAEIN     STANDARD;      PRT;  1541 AA.
AC   P42782;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK368 / SEROTYPE B;
RX   MEDLINE=89379374; PubMed=2506130;
RA   Poulsen K., Brandt J., Hjorth J.P., Thoegersen H.C., Kilian M.;
RT   "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RT   of Haemophilus influenzae serotype b.";
RL   Infect. Immun. 57:3097-3105(1989).
RN   [2]
RP   MUTAGENESIS OF SER-288.
RC   STRAIN=HK368 / SEROTYPE B;
RX   MEDLINE=92234949; PubMed=1373717;
RA   Poulsen K., Reinholdt J., Kilian M.;
RT   "A comparative genetic study of serologically distinct Haemophilus
RT   influenzae type 1 immunoglobulin A1 proteases.";
RL   J. Bacteriol. 174:2913-2921(1992).
CC   -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC       PRODUCING INTACT FC AND FAB FRAGMENTS.
CC   -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC       certain Pro-|-Xaa bonds in the hinge region. No small molecule
CC       substrates are known.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC       SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC       OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC       DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64357; CAA45708.1; -.
DR   EMBL; M87492; AAA24969.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26   1008       IMMUNOGLOBULIN A1 PROTEASE.
FT   PROPEP    1009   1541       HELPER PEPTIDE (POTENTIAL).
FT   ACT_SITE   288    288       PROBABLE.
FT   MUTAGEN    288    288       S->T: LOSS OF ACTIVITY.
SQ   SEQUENCE  1541 AA;  169370 MW;  CE7257CB3196C600 CRC64;


  Query Match             5.7%;  Score 95;  DB 1;  Length 1541;
  Best Local Similarity   23.5%;  Pred. No. 14;
  Matches   65;  Conservative   39;  Mismatches  101;  Indels   72;  Gaps   17;

Qy      17 FNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTAQFYEIKPT 76
           |   |:: |   : |   : | |   || |  :|      ||:  : | :      | |
Db      37 FRDFAENKGKFSV-GATNVLVKDKNNKDLGTALP------NGIPMIDFSVVDVDKRI-AT 88

Qy      77 KENEQYIGMLRQAVKNESPVHI-FLKPNSNEIGKVES---ASPEDVRYF---------- 121
             | ||:  ::      | :|   |  | | |  |  ::    | |: |||
Db      89 LINPQYVVGVKHVSNGVSELHFGNLNGNMNN-GNAKAHRDVSSEENRYFSVEKNEYPTKL 147

Qy     122 --KTILTKEVKGQTNK---------LASVIPDVATLNSLFNQIKNQSCGTSTASSPCITF 170
             ||: |::   || |          |   : :|| : :            |||||   |:
Db     148 NGKTVTTED---QTQKRREDYYMPRLDKFVTEVAPIEA-----------STASSDAGTY 192

Qy     171 ----RYPVDGCYARAHKMRQILMNNGYD----CEKQFVYGNLKASTGTCCVAWSYHVA-- 220
```

```
              :||    : |     | :   | :         | ||     |    |::| :|
Db    193 NDQNKYP---AFVRLGSGSQFIYKKGDNYSLILNNHEVGGNNLKLVGD---AYTYGIAGT 246

Qy    221 -ILVSYKN----ASGVTEKRIIDP-SLFSSGPVTDTA 251
          |:::|       | :::   ||  : |  |:|: |
Db    247 PYKVNHENNGLIGFGNSKEEHSDPKGILSQDPLTNYA 283


RESULT   7
YF54_METJA
ID   YF54_METJA     STANDARD;      PRT;   622 AA.
AC   Q58949;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ1554 precursor.
GN   MJ1554.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: TO A.FULGIDUS AF0817.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67596; AAB99574.1; -.
DR   TIGR; MJ1554; -.
KW   Hypothetical protein; Signal; Complete proteome.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    622       HYPOTHETICAL PROTEIN MJ1554.
SQ   SEQUENCE   622 AA;  72282 MW;  86CAA62F3854D73F CRC64;


  Query Match             5.6%;  Score 94.5;  DB 1;  Length 622;
  Best Local Similarity   20.9%;  Pred. No. 4.9;
  Matches   73;  Conservative   47;  Mismatches  107;  Indels  123;  Gaps   20;

Qy      4 LFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLK----DFGKTVPVGIDEENGM 59
          :|||::  :::  |: | :               |:  :|    || | :||
Db      8 IFLSLLMIISL--FSGCVE--------------NEKPIKEGSNDF-KLIPVN------- 42

Qy     60 IKVSFMLTAQFYEIKPTKENE--QYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPED 117
                : | | | ||    || :      |  :  :| :: |       | |
Db     43 ------SKSNFEEFKNTVENSIGNYIYVGHSYASREVQITSTVKSSNVE------TSTEP 90

Qy    118 VRYFKTILTKEVKG-------QTN-------------------KLASVIPDVA------- 144
           |: ||    :|||        :||                  | | :| :::
Db     91 ERFSKT--NVQVKGVDEADILKTNGNIIAFSQNKIYLIKPLPPKYAKIIKNISECGYLYL 148

Qy    145 TLNSL----FNQIKNQSCGTSTASSPCITFRYPVDGCYARAH----KMRQILMNNGYDCE 196
          | |:|    :|:|   |   |    | | ::  ::| |  :      :  ::  |  ||
Db    149 TNNTLIVISWNKI--TSYNVSNPEMPKIIWQMDLNGSYVDSRLYNGTLYLVVRKNSIDC- 205

Qy    197 KQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNAC 256
             |: | |           |::  |                 | ::|     || :  :
Db    206 -PIVWNNYKIGYD------KYYIPEL---------------PPIYSMD--FDTTYIISR 240

Qy    257 VNTSCGS-----ASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTKF 301
          :|    |      | | :|  |     | | :| |   |  ||   ::  |
Db    241 INIKSGKVENSIAIVGNYKTT----LYMSKNNLYFAYNLKINEKKLMLNF 286


RESULT   8
D7_DICDI
ID   D7_DICDI       STANDARD;      PRT;   850 AA.
AC   P54682;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   cAMP-inducible prespore protein D7 precursor.
GN   D7.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=95080502; PubMed=7988791;
RA   Agarwal A., Sloger M.S., Oyama M., Blumberg D.D.;
RT   "Analysis of a novel cyclic Amp inducible prespore gene in
RT   Dictyostelium discoideum: evidence for different patterns of cAMP
RT   regulation.";
RL   Differentiation 57:151-162(1994).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PRESPORE CELLS.
CC   -!- INDUCTION: BY CAMP.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U25143; AAA73514.1; -.
DR   DictyDb; DD02038; -.
KW   Sporulation; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    850       CAMP-INDUCIBLE PRESPORE PROTEIN D7.
FT   DOMAIN      470    475       POLY-GLN.
FT   DOMAIN      555    568       POLY-ASN.
FT   DOMAIN      728    738       POLY-GLN.
SQ   SEQUENCE   850 AA;  95343 MW;  13BA634CCE7AA502 CRC64;


  Query Match             5.6%;  Score 94.5;  DB 1;  Length 850;
  Best Local Similarity   23.1%;  Pred. No. 7.3;
  Matches   74;  Conservative   42;  Mismatches  111;  Indels   93;  Gaps   15;

Qy     10 AFVTVLTFN---SCADSNGNQEINGK---EKLSVNDSKLKDFGKTVPVGIDEENGMIKVS 63
             :|| ||  :   |   ||   |  ||   :||: |  |::|      | :  ||   :|
Db      8 SFVIVLILSCIISTCTSNQISEDVGKAINQKLNENIEKVED------VVVQFENVANQV- 60

Qy     64 FMLTAQFYEIKPTKENEQYIG-----MLRQAVKNE-----SPVHIFLKPNSNEIGKVESA 113
                |:|   : :: :|      || :  |        :  :|  ::|:: ||| |
Db     61 ------IEELKVEHQRQELLGEVKHDSLRDSATNYIWGLLDKIQSYLPKDNNKVSKVEEA 114

Qy    114 SPEDVRYFKTILTKEVKGQTNKLASVIPDV--ATLNSLFNQIK-----NQSCGTST---- 162
                        || | : | | |   |: :  |  |      :|| |:||
Db    115 FS-------------SGQNNNIGSSIGDSTGASTSPQFQSINGLSGASQSSGSSTGGTG 160

Qy    163 -----ASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKAS-TGTCCVAWS 216
                ::  | |   |     :   : |: :            :| |: ||   |: :
Db    161 DSDSKTTNEAIIFSSKVSTTDRQESIIGQVAIT---------AKDSLGATITGLGGVSST 211
```

```
Qy     217 YHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSASV---------- 266
           |   ::   | |          : : |  | |  | |   |  : :|
Db     212 AKVGGQITNGRAQG---------QVITGGDNTGTVGRGAVTTASAVANTVGEFLGGSRTG 262

Qy     267 -SSYANTAGNVYYRSPSNSY 285
           || | | |||     |:||
Db     263 GSSSAGTVGNVI----SDSY 278


RESULT   9
CNA_STAAU
ID   CNA_STAAU      STANDARD;      PRT;  1183 AA.
AC   Q53654;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen adhesin precursor.
GN   CNA.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FDA 574;
RX   MEDLINE=92165839; PubMed=1311320;
RA   Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA   Lindberg M., Hoeoek M.;
RT   "Molecular characterization and expression of a gene encoding a
RT   Staphylococcus aureus collagen adhesin.";
RL   J. Biol. Chem. 267:4766-4772(1992).
RN   [2]
RP   ERRATUM.
RA   Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA   Lindberg M., Hoeoek M.;
RL   J. Biol. Chem. 269:11672-11672(1994).
RN   [3]
RP   COLLAGEN-BINDING DOMAIN.
RC   STRAIN=FDA 574;
RX   MEDLINE=94032261; PubMed=8218209;
RA   Patti J.M., Boles J.O., Hoeoek M.;
RT   "Identification and biochemical characterization of the ligand
RT   binding domain of the collagen adhesin from Staphylococcus aureus.";
RL   Biochemistry 32:11428-11435(1993).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX   MEDLINE=97475225; PubMed=9334749;
RA   Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA   Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,
RA   Narayana S.V.L.;
RT   "Structure of the collagen-binding domain from a Staphylococcus
RT   aureus adhesin.";
RL   Nat. Struct. Biol. 4:833-838(1997).
CC   -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYOCOCCAL CELLS TO
CC       COLLAGEN-CONTAINING SUBSTRATA.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC   -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC       IN THE REGION OF THE MEMBRANE ANCHOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81736; AAA20874.1; -.
DR   PDB; 1AMX; 24-JUN-98.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW   Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30   1183       COLLAGEN ADHESIN.
FT   DOMAIN       30   1157       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1158   1177       MEMBRANE ANCHOR (POTENTIAL).
FT   DOMAIN     1178   1183       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      151    318       COLLAGEN-BINDING.
FT   DOMAIN      533   1093       3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT   DOMAIN     1093   1157       LYS/PRO-RICH (CELL WALL-SPANNING).
FT   DOMAIN     1151   1156       CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT                                PROTEINS.
FT   REPEAT      533    719       B1.
FT   REPEAT      720    906       B2.
FT   REPEAT      907   1093       B3.
SQ   SEQUENCE   1183 AA;  133066 MW;  B6A1CC072E575D76 CRC64;


  Query Match             5.5%;  Score 93;  DB 1;  Length 1183;
  Best Local Similarity   21.4%;  Pred. No. 15;
  Matches   73;  Conservative   39;  Mismatches  127;  Indels  102;  Gaps   16;

Qy       2 KNLFLSMMAFVTVLT-----FNSCADSNGNQEINGKE-------KLSVNDSKLKDFGK-T 48
           ||: |  | |: :|      ||       |:    ::        |:|: ||::| || |
Db       3 KNV-LKFMVFIMLLNIITPLFNK------NEAFAARDISSTNVTDLTVSPSKIEDGGKTT 55

Qy      49 VPVGIDEENG------MIKVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKP 102
           | :  |::||      ||||::  :         | : | |   :   || |      : |
Db      56 VKMTFDDKNGKIQNGDMIKVAWPTSG-------TVKIEGYSKTVPLTVKGEQVGQAVITP 108

Qy     103 NSNEI---GKVESASPEDVRYFKTILTKEVKG----QTNKLASVIPDVATLNSLFN-QIK 154
           :   |    |||  |  ||  |      ||:|    |||     :  : :  |   | :
Db     109 DGATITFNDKVEKLS--DVSGFAEF---EVQGRNLTQTNTSDDKVATITSGNKSTNVTVH 163

Qy     155 NQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVA 214
             |||:                                      | |      ||
Db     164 KSEAGTSSV----------------------------------FYY-----KTGDMLPE 183

Qy     215 WSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACV-----NTSCGSASVSSY 269
           : ||   ::  |      | |        |   | :  |  |     |   | :::: :
Db     184 DTTHVRWFLNINNEKSYVSKDITIKDQIQGGQQLDLSTLNINVTGTHSNYYSGQSAITDF 243

Qy     270 A----------NTAGNVYYRSPSNSYLYDNNLINTNCVLT 299
                    ||   :    |    |::  ||    :|
Db     244 EKAFPGSKITVDNTKNTIDVTIPQGYGSYNSFSINYKTKIT 284


RESULT  10
CYS5_DICDI
ID   CYS5_DICDI     STANDARD;      PRT;   344 AA.
AC   P54640;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Cysteine proteinase 5 precursor (EC 3.4.22.-).
GN   CPRE OR CP5.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=96081966; PubMed=7499424;
RA   Souza G.M., Hirai J., Mehta D.P., Freeze H.H.;
RT   "Identification of two novel Dictyostelium discoideum cysteine
RT   proteinases that carry N-acetylglucosamine-1-P-modification.";
RL   J. Biol. Chem. 270:28938-28945(1995).
CC   -!- SUBCELLULAR LOCATION: Lysosomal.
CC   -!- DEVELOPMENTAL STAGE: PRESENT IN THE VEGETATIVE PHASE AND DECREASES
CC       WITH THE START DEVELOPMENT, REAPPEARS IN LOW LEVELS WHEN THE
CC       FRUITING BODY IS FORMED.
CC   -!- PTM: PHOSPHOGLYCOSYLATED, CONTAINS GLCNAC-ALPHA-1-P-SER RESIDUES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC       PAPAIN FAMILY OF THIOL PROTEASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L36205; AAA92018.1; -.
DR   HSSP; P07711; 1CJL.
DR   MEROPS; C01.081; -.
DR   DictyDb; DD01061; cprE.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Thiolprot_act_site.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW   Hydrolase; Thiol protease; Lysosome; Zymogen; Glycoprotein;
KW   Phosphorylation; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   PROPEP       18    111       ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN       112    344       CYSTEINE PROTEINASE 5.
FT   DOMAIN      196    340       SER-RICH.
FT   ACT_SITE    136    136       BY SIMILARITY.
FT   ACT_SITE    272    272       BY SIMILARITY.
FT   ACT_SITE    311    311       BY SIMILARITY.
FT   DISULFID    133    174       BY SIMILARITY.
FT   DISULFID    167    207       BY SIMILARITY.
FT   DISULFID    265    333       BY SIMILARITY.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    297    297       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   344 AA;  37212 MW;  82F3F35B8C147BA8 CRC64;
```

```
  Query Match             5.5%;  Score 92;  DB 1;  Length 344;
  Best Local Similarity   22.7%;  Pred. No. 3.7;
  Matches   76;  Conservative   36;  Mismatches  117;  Indels  106;  Gaps   16;

Qy      55 EENGMIKVSFMLTAQFYEIKPTKENEQYIGMLRQA-VKNESPVHIFLKPNSNEIGKVESA 113
           || |     |     : :   :| :|  :|:   | : ||   : :|        | :::
Db      44 EEFGARYNIFTANMDYVQQWNSKGSETVLGLNNFADITNEEYRNTYLGT------KFDAS 97

Qy     114 SPEDVRYFKTILTKEVKGQTNKLA---------SVIPDVATLNSLFNQIKNQ-SCG---- 159
           |         | |:| |  ||  |          :| |           :|||  ||
Db      98 S--------LIGTQEEKVHTNSSAASKDWRSEGAVTP----------VKNQGQCGGCWS 138

Qy     160 -TSTASS-------------------PCITFRYPVDG---CYARAHKMRQILMNNGYDC 195
            ::| |:                   | |     ||    ||       |: ||| |
Db     139 FSTTGSTEGAHFQSKGELVSLSEQNLIDCSTENSGCDGGLMTYA----FEYIINNNGIDT 194

Qy     196 EKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRI-----------IDPS--- 240
           |  : |   ||  | |        | | |||  :  :|  :            || |
Db     195 ESSYPY---KAENGKCEYKSENSGATLSSYKTVTAGSESSLESAVNVNPVSVAIDASHQS 251

Qy     241 --LFSSGPVTDTAWRNAC-------------VNTSCGSASVSSYANTAGNVYYRSPSNSY 285
             |::||      :   |               :  ||:|  |   ::||:   | || |
Db     252 FQLYTSG----IYYEPECSSENLDHGVLAVGYGSGSGSSSGQSSGQSSGNL-SASSSNEY 306

Qy     286 LYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCGF 320
           |   |:  :  : |:|    :     ||  |
Db     307 WIVKNSWGTSWGIEGYILMSRNRDNNCGIASSASF 341
```

```
RESULT  11
NUD1_YEAST
ID   NUD1_YEAST     STANDARD;      PRT;   851 AA.
AC   P32336; Q08895;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NUD1 protein.
GN   NUD1 OR YOR373W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dulic V., Zanolari B., Riezman H.;
RL   Submitted (SEP-1991) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Delius H., Hebling U., Hofmann B.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR NUCLEAR DIVISION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62147; CAA44073.1; -.
DR   EMBL; Z75281; CAA99704.1; -.
DR   PIR; S19056; S19056.
DR   SGD; S0005900; NUD1.
DR   InterPro; IPR001611; LRR.
DR   Pfam; PF00560; LRR; 5.
DR   PRINTS; PR00019; LEURICHRPT.
KW   Cell cycle.
FT   DOMAIN      223    239       POLY-ASN.
FT   DOMAIN      266    274       POLY-SER.
FT   CONFLICT      5      6       TQ -> SE (IN REF. 2).
SQ   SEQUENCE   851 AA;  94103 MW;  CB9F0408633C1315 CRC64;
```

```
  Query Match             5.4%;  Score 91.5;  DB 1;  Length 851;
  Best Local Similarity   21.2%;  Pred. No. 13;
  Matches   34;  Conservative   26;  Mismatches   57;  Indels   43;  Gaps    6;

Qy      24 NGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTAQFYEIKPTKENEQYI 83
           |  :::    |    |    |   || | :  ||  |   :|  : ||:  ||  :  :
Db     314 NQGLDVSSSHSLDNTSSNQSQFATMVPTGDNHTNG--KAPSILDKKAYELTSTKPGD--V 369

Qy      84 GMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTK-----EVKGQTNKLAS 138
           |  :: :: |  :      ||::       :| |   |  : ||     :||||     |
Db     370 GYRQKKIQEEENL-----ANSDD-------TPLDTPKFNDLFTKNGTRAKVKGQMRTSRS 417

Qy     139 V----------------------IPDVATLNSLFNQIKNQ 156
           :                      |  :: :|: ||: : |
Db     418 ISNSNLLEAHKKLKTFPAERVEDITSISEVNTSFNETEKQ 457
```

```
RESULT  12
DPOZ_HUMAN
ID   DPOZ_HUMAN     STANDARD;      PRT;  3130 AA.
AC   O60673; O43214;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   DNA polymerase zeta catalytic subunit (EC 2.7.7.7) (hREV3).
GN   REV3L OR POLZ OR REV3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=98284025; PubMed=9618506;
RA   Gibbs P.E.M., McGregor W.G., Maher V.M., Nisson P., Lawrence C.W.;
RT   "A human homolog of the Saccharomyces cerevisiae REV3 gene, which
RT   encodes the catalytic subunit of DNA polymerase zeta.";
```

```
RL   Proc. Natl. Acad. Sci. U.S.A. 95:6876-6880(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow, and Leukocyte;
RX   MEDLINE=99202263; PubMed=10102035;
RA   Lin W., Wu X., Wang Z.;
RT   "A full-length cDNA of hREV3 is predicted to encode DNA polymerase
RT   zeta for damage-induced mutagenesis in humans.";
RL   Mutat. Res. 433:89-98(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Murakumo Y., Rasio D., Roth T., Negrini M., Croce C.M., Fishel R.;
RT   "Cloning and characterization of hREV3, the human homolog of S.
RT   cerevisiae REV3.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 79-3130 FROM N.A.
RA   Roth T., Rasio D., Murakumo Y., Negrini M., Croce C.M., Fishel R.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 79-3130 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=99126315; PubMed=9925914;
RA   Morelli C., Mungall A.J., Negrini M., Barbanti-Brodano G., Croce C.M.;
RT   "Alternative splicing, genomic structure, and fine chromosome
RT   localization of REV3L.";
RL   Cytogenet. Cell Genet. 83:18-20(1998).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC   -!- DOMAIN: ITS C-TERMINAL PART COULD SERVE AS THE CATALYTIC DOMAIN
CC       DURING NUCLEOTIDE POLYMERIZATION, WHILE ITS N-TERMINAL PART COULD
CC       PROVIDE SITES FOR PROTEIN-PROTEIN INTERACTIONS WITH OTHER FACTORS
CC       DURING TRANSLESION DNA SYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF058701; AAC24357.1; -.
DR   EMBL; AF071798; AAC24009.1; -.
DR   EMBL; AF157476; AAD40184.1; -.
DR   EMBL; AF179428; AAG09402.1; -.
DR   EMBL; AF179429; AAG09403.1; -.
DR   EMBL; AF035537; AAB88486.1; -.
DR   EMBL; AF078695; AAC28460.1; -.
DR   MIM; 602776; -.
DR   InterPro; IPR002064; DNA_pol_B.
DR   Pfam; PF00136; DNA_pol_B; 1.
DR   Pfam; PF03104; DNA_pol_B_exo; 2.
DR   PRINTS; PR00106; DNAPOLB.
DR   SMART; SM00486; POLBc; 1.
DR   PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   DNA-binding; DNA repair; Nuclear protein; Zinc-finger; Polymorphism.
FT   ZN_FING    3042   3057       C4-TYPE (POTENTIAL).
FT   ZN_FING    3086   3104       C4-TYPE (POTENTIAL).
FT   VARIANT     231    231       Q -> H.
FT                                /FTId=VAR_008516.
FT   VARIANT     389    389       S -> T.
FT                                /FTId=VAR_008517.
FT   VARIANT    1540   1540       K -> E.
FT                                /FTId=VAR_008518.
FT   VARIANT    2607   2607       S -> T.
FT                                /FTId=VAR_008519.
FT   CONFLICT    237    237       E -> Q (IN REF. 4 AND 5).
FT   CONFLICT   1156   1156       Y -> C (IN REF. 4 AND 5).
SQ   SEQUENCE   3130 AA;  352782 MW;  1C0700900F10BB14 CRC64;


  Query Match             5.4%;  Score 91.5;  DB 1;  Length 3130;
  Best Local Similarity   19.0%;  Pred. No. 67;
  Matches   60;  Conservative   51;  Mismatches  112;  Indels   93;  Gaps   11;

Qy      2 KNLFLSMMAFVTVLTF---------NSCADSNGNQEINGKEKLSVN--DSKLKDFGKTVP 50
          | |  |: :|    |:          ::    | ::| : || | |:  :| : |: : :|
Db    608 KGLDNSVTSFTNESTYSMKYPGSLSSTVHSENSHKENSKKEILPVSSCESSIFDYEEDIP 667

Qy     51 VGIDEENGMIKVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKV 110
                 :| |    | |         :|:  |:   :|:   || | :||
Db    668 -------------SVTRQVPSRKYTN--------IRKIEKDSPFIHMHRHPNENTLGK- 704

Qy    111 ESASPEDVRYFKTILTKE--VKGQTNKLASVIP-------------------------- 141
          | :  |: : |  :: |   || :  |:|: |
Db    705 NSFNFSDLNHSKNKVSSEGNEKGNSTALSSLFPSSFTENCELLSCSGENRTMVHSLNSTA 764

Qy    142 DVATLNSL------FNQIKNQSCGTSTASSPCITFRYPV-DGCYARAHKMRQILMN---- 190
          | : || |      | : | :    |  ::  : |   |   |  |  |:  :
Db    765 DESGLNKLKIRYEEFQEHKTEKPSLSQQAAHYMFFPSVVLSNCLTRPQKLSPVTYKLQPG 824

Qy    191 ---NGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPV 247
             :    |: : |: : || :                  : :| |:   |  :  :| |
Db    825 NKPSRLKLNKRKLAGHQETSTKS----------------SETGSTKDNFIQNNPCNSNPE 868

Qy    248 TDTAWRNACVNTSCGS 263
          | |  :    |: |:
Db    869 KDNALASDLTKTTRGA 884


RESULT  13
NCA2_HUMAN
ID   NCA2_HUMAN     STANDARD;      PRT;   761 AA.
AC   P13592; P13593;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, phosphatidylinositol-linked isoform
DE   precursor (N-CAM 120) (NCAM-120) (CD56 antigen).
GN   NCAM1 OR NCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (N-CAM 120).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=89305258; PubMed=3253057;
RA   Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,
RA   Elsom V., Moore S.E., Goridis C., Walsh F.S.;
RT   "Complete sequence and in vitro expression of a tissue-specific
RT   phosphatidylinositol-linked N-CAM isoform from skeletal muscle.";
RL   Development 104:165-173(1988).
RN   [2]
RP   SEQUENCE OF 491-761 FROM N.A. (N-CAM 120).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=87301755; PubMed=2887295;
RA   Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA   Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT   "Human muscle neural cell adhesion molecule (N-CAM): identification
RT   of a muscle-specific sequence in the extracellular domain.";
RL   Cell 50:1119-1130(1987).
RN   [3]
RP   SEQUENCE OF 491-655 FROM N.A. (SECRETED ISOFORM).
RX   MEDLINE=89077552; PubMed=3203385;
RA   Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,
RA   Dickson G., Walsh F.S.;
RT   "Alternative splicing generates a secreted form of N-CAM in muscle
RT   and brain.";
RL   Cell 55:955-964(1988).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
```

```
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16841; CAA34739.1; -.
DR   EMBL; M17409; AAA59912.1; -.
DR   EMBL; M22094; AAA59910.1; -.
DR   EMBL; M22092; AAA59911.1; -.
DR   EMBL; M22091; AAA59911.1; JOINED.
DR   PIR; S07784; IJHUNG.
DR   PIR; A31635; A31635.
DR   MIM; 116930; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW   GPI-anchor; Alternative splicing.
FT   SIGNAL        1     19
FT   CHAIN        20    761       NEURAL CELL ADHESION MOLECULE,
FT                                 PHOSPHATIDYLINOSITOL-LINKED ISOFORM.
FT   DOMAIN        34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN       132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN       228    294       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN       322    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN       419    486       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN       518    595       FIBRONECTIN TYPE-III 1.
FT   DOMAIN       660    727       FIBRONECTIN TYPE-III 2.
FT   DISULFID      41     96       PROBABLE.
FT   DISULFID     139    189       PROBABLE.
FT   DISULFID     235    287       PROBABLE.
FT   DISULFID     329    385       PROBABLE.
FT   DISULFID     426    479       PROBABLE.
FT   CARBOHYD     222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     635    655       GEPSAPKLEGQMGEDGNSIKV -> NIAQNHCCNMFQAGLH
FT                                 NALMK (IN SECRETED ISOFORM).
SQ   SEQUENCE   761 AA;  83770 MW;  F0CAD3292D7AB67E CRC64;


  Query Match             5.4%;  Score 91;  DB 1;  Length 761;
  Best Local Similarity  20.6%;  Pred. No. 12;
  Matches   67;  Conservative   45;  Mismatches  115;  Indels   98;  Gaps   16;

Qy      44 DFGKTVPVGIDEENGMIKVSFMLTAQFYEIKPTKENEQYIG-------MLRQAVKNESPV 96
           : |::| :  | | |  : :   |    :|:  :::|:||         :::  ||:
Db     226 NLGQSVTLVCDAE-GFPEPTMSWTKDGEQIEQEEDDEKYIFSDDSSQLTIKKVDKNDEAE 284

Qy      97 HIFLKPNSNEIGKVESASPEDVR-YFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKN 155
             :| :  |         | :|  : |     ::    |: |  : :  ||
Db     285 YICIAEN--------KAGEQDATIHLKVFAKPKITYVENQTAMELEEQVTL--------- 327

Qy     156 QSCGTSTASSPCITFR----------YPVDG-CYARAH--------KMRQILMNNGYDCE 196
           :|  |    | ||:|            :||    |:|        |  |      | |
Db     328 -TCEASGDPIPSITWRTSTRNISSEEKTLDGHMVVRSHARVSSLTLKSIQYTDAGEYICT 386

Qy     197 KQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNAC 256
           |    |           : : | |             | |   |||    |
Db     387 ASNTIGQDSQS----------MYLEVQYA-----------PKL--QGPVAVYTWEGNQ 421

Qy     257 VNTSC-----GSASVSSY--------ANTAGNVYYRSPSNSYL-----YDNNLINTNCVL 298
           || :|      ||::| :        :| :    | :|| |||      :|:  | ||
Db     422 VNITCEVFAYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTA 481

Qy     299 T--------KFSLLSGCSPSPAPDV 315
                   :| |:   :|| :| :
Db     482 VNRIGQESLEFILVQADTPS-SPSI 505


RESULT  14
NCA1_HUMAN
ID   NCA1_HUMAN     STANDARD;      PRT;   848 AA.
AC   P13591; Q16180; Q15829;
DT   01-JAN-1990 (Rel. 13, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE   (NCAM-140) (CD56 antigen).
GN   NCAM1 OR NCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94356433; PubMed=8075973;
RA   Saito S., Tanio Y., Tachibana I., Hayashi S., Kishimoto T., Kawase I.;
RT   "Complementary DNA sequence encoding the major neural cell adhesion
RT   molecule isoform in a human small cell lung cancer cell line.";
RL   Lung Cancer 10:307-318(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91250739; PubMed=1710251;
RA   Lanier L.L., Chang C., Azuma M., Ruitenberg J.J., Hemperly J.J.,
RA   Phillips J.H.;
RT   "Molecular and functional analysis of human natural killer cell-
RT   associated neural cell adhesion molecule (N-CAM/CD56).";
RL   J. Immunol. 146:4421-4426(1991).
RN   [3]
RP   SEQUENCE OF 491-848 FROM N.A.
RX   MEDLINE=87301755; PubMed=2887295;
RA   Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA   Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT   "Human muscle neural cell adhesion molecule (N-CAM): identification
RT   of a muscle-specific sequence in the extracellular domain.";
RL   Cell 50:1119-1130(1987).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S71824; AAB31836.1; -.
DR   EMBL; U63041; AAB04558.1; -.
DR   EMBL; M17410; AAA59913.1; -.
DR   PIR; B26883; B26883.
DR   HSSP; P40189; 1BQU.
DR   MIM; 116930; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW   Transmembrane; Alternative splicing.
FT   SIGNAL        1     19
FT   CHAIN        20    848       NEURAL CELL ADHESION MOLECULE, 140 KDA
FT                                 ISOFORM.
FT   DOMAIN       20    708       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    709    729       POTENTIAL.
FT   DOMAIN      730    848       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    294       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      322    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      419    486       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    595       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      660    727       FIBRONECTIN TYPE-III 2.
FT   DISULFID     41     96       PROBABLE.
FT   DISULFID    139    189       PROBABLE.
FT   DISULFID    235    287       PROBABLE.
FT   DISULFID    329    385       PROBABLE.
FT   DISULFID    426    479       PROBABLE.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    215    215       Q -> R (IN REF. 2).
FT   CONFLICT    239    239       G -> R (IN REF. 2).
FT   CONFLICT    490    490       L -> F (IN REF. 2).
FT   CONFLICT    599    600       QG -> R (IN REF. 3).
FT   CONFLICT    720    721       MISSING (IN REF. 3).
FT   CONFLICT    811    811 .     G -> A (IN REF. 3).
SQ   SEQUENCE   848 AA;  93360 MW;  68D2F0C0E6C1C2AD CRC64;


  Query Match             5.4%;  Score 91;  DB 1;  Length 848;
  Best Local Similarity   20.6%;  Pred. No. 14;
  Matches   67;  Conservative   45;  Mismatches  115;  Indels   98;  Gaps   16;

Qy      44 DFGKTVPVGIDEENGMIKVSFMLTAQFYEIKPTKENEQYIG-------MLRQAVKNESPV 96
           : |::| :  | | |  : :   |    :|:  :::|:||         :::  ||:
Db     226 NLGQSVTLVCDAE-GFPEPTMSWTKDGEQIEQEEDDEKYIFSDDSSQLTIKKVDKNDEAE 284

Qy      97 HIFLKPNSNEIGKVESASPEDVR-YFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKN 155
           :| :  |         |  :|   : |     ::    |: |  : :  ||
Db     285 YICIAEN--------KAGEQDATIHLKVFAKPKITYVENQTAMELEEQVTL--------- 327

Qy     156 QSCGTSTASSPCITFR----------YPVDG-CYARAH--------KMRQILMNNGYDCE 196
           :|  |    | ||:|            :||    |:|        |  |      | |
Db     328 -TCEASGDPIPSITWRTSTRNISSEEKTLDGHMVVRSHARVSSLTLKSIQYTDAGEYICT 386

Qy     197 KQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNAC 256
           |   |          : : | |             | |   |||    |
Db     387 ASNTIGQDSQS----------MYLEVQYA-----------PKL--QGPVAVYTWEGNQ 421

Qy     257 VNTSC-----GSASVSSY--------ANTAGNVYYRSPSNSYL-----YDNNLINTNCVL 298
           || :|      ||::| :        :| :    | :|| |||      :|:  | ||
Db     422 VNITCEVFAYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTA 481

Qy     299 T--------KFSLLSGCSPSPAPDV 315
                   :| |:   :|| :| :
Db     482 VNRIGQESLEFILVQADTPS-SPSI 505


RESULT  15
YLK3_CAEEL
ID   YLK3_CAEEL     STANDARD;      PRT;  1895 AA.
AC   P41951;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Putative serine/threonine-protein kinase D1044.3 in chromosome III
DE   (EC 2.7.1.-).
GN   D1044.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Pauley A., Waterston R.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       STRONG, TO ZC84.1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00065; AAA50735.1; -.
DR   HSSP; Q63450; 1A06.
DR   WormPep; D1044.3; CE01206.
DR   InterPro; IPR002899; EB.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF01683; EB; 12.
DR   Pfam; PF00069; pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00289; WR1; 12.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW   ATP-binding.
FT   DOMAIN      431    703       PROTEIN KINASE.
FT   NP_BIND     437    445       ATP (BY SIMILARITY).
FT   BINDING     461    461       ATP (BY SIMILARITY).
FT   ACT_SITE    569    569       BY SIMILARITY.
SQ   SEQUENCE   1895 AA;  208393 MW;  F23C9F7881353AD6 CRC64;


  Query Match             5.3%;  Score 89;  DB 1;  Length 1895;
  Best Local Similarity   24.1%;  Pred. No. 56;
  Matches   46;  Conservative   27;  Mismatches   88;  Indels   30;  Gaps    9;

Qy     155 NQSC----------GTSTASSPCITFRYPVDGCYARAHKM-RQILMNNGYDCEKQF--VY 201
           || |          || : :| |||   |   |   :  :   : ||    |   :  ||
Db    1112 NQQCVISNGLNCQLGTVSYNSQCITLASPGQNCQTSSQCIDNSVCMNQMCTCNNNYRLVY 1171

Qy     202 GNLKASTGTCCVAWSYHV---AILVSYKNASGVTEKRIIDPSLFSSGPVT----DTAWRN 254
           |     | : |      |    :|:|    : :  :: :  :: :||       ||
Db    1172 GYCVPITSSICQQTQTLVNNQCVLLSIVGETCIANQQCVGGAMCNSGTCQCTNGATAMYG 1231

Qy     255 ACVNTSCGSASVSSYANTAGNVYYR-----SPSNSYLYDNNLINTN--CVLTKFSLLSGC 307
           |:::|  | : |:  :  | |       | | |    || : ||  || |  |:   |
Db    1232 YCISSSSSSCN-SNQVSINGMCYNTVQVGGSCSFSQQCLNNAVCTNNICVSTFCSV--SC 1288
```

```
Qy     308 SPSPAPDVSSC 318
           |  :     :  |
Db    1289 STNQVCISNQC 1299
```

Search completed: June 27, 2002, 21:59:22
Job time: 665 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:58:26 ; Search time 102.74 Seconds
(without alignments)
538.820 Million cell updates/sec

Title: US-09-727-769A-8
Perfect score: 1679
Sequence: 1 MKNLFLSMMAFVTVLTFNSC.....FSLLSGCSPSPAPDVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1679	100.0	320	2	Q9AQQ8	Q9aqq8 chryseobact
2	108.5	6.5	342	13	P70004	P70004 xenopus lae
3	108	6.4	927	5	O02364	O02364 caenorhabdi
4	102	6.1	2673	4	Q96SC3	Q96sc3 homo sapien
5	102	6.1	5636	4	Q96RW7	Q96rw7 homo sapien
6	101.5	6.0	551	16	Q9HYT4	Q9hyt4 pseudomonas
7	101	6.0	2395	5	Q27167	Q27167 paramecium
8	100	6.0	848	15	O40366	O40366 human immun
9	99.5	5.9	534	3	O94317	O94317 schizosacch
10	99	5.9	594	10	Q9S746	Q9s746 arabidopsis
11	98.5	5.9	419	16	Q9KB36	Q9kb36 bacillus ha
12	97	5.8	534	17	O26459	O26459 methanother
13	97	5.8	564	12	O72736	O72736 cowpox viru
14	96.5	5.7	260	16	Q99RZ3	Q99rz3 staphylococ
15	96.5	5.7	5458	5	Q9U459	Q9u459 plasmodium
16	96	5.7	423	2	Q9F961	Q9f961 anaplasma m
17	95.5	5.7	673	16	Q97TK7	Q97tk7 clostridium
18	95	5.7	423	2	Q9Z620	Q9z620 anaplasma m
19	93.5	5.6	1894	11	P70206	P70206 mus musculu
20	93	5.5	347	2	Q93R80	Q93r80 porphyromon
21	93	5.5	594	10	Q9SXZ3	Q9sxz3 arabidopsis
22	92	5.5	1154	10	Q9LPQ6	Q9lpq6 arabidopsis
23	91.5	5.4	427	17	Q971P3	Q971p3 sulfolobus
24	91.5	5.4	2013	16	Q92EK2	Q92ek2 listeria in
25	91.5	5.4	2909	4	Q9NU25	Q9nu25 homo sapien
26	91	5.4	392	10	Q9FZJ9	Q9fzj9 arabidopsis
27	91	5.4	396	10	Q9ZW53	Q9zw53 arabidopsis
28	91	5.4	876	5	Q9XZN5	Q9xzn5 mya arenari
29	90.5	5.4	843	5	O02264	O02264 caenorhabdi
30	90.5	5.4	1166	12	Q82666	Q82666 avian infec
31	90	5.4	243	5	Q18804	Q18804 caenorhabdi
32	89.5	5.3	1124	10	Q9SF08	Q9sf08 arabidopsis
33	89.5	5.3	1578	16	Q92E25	Q92e25 listeria in
34	89.5	5.3	1836	10	Q9LRX4	Q9lrx4 arabidopsis
35	89.5	5.3	2500	5	O96223	O96223 plasmodium
36	89.5	5.3	2510	5	Q94658	Q94658 plasmodium
37	89	5.3	595	9	Q9AZR8	Q9azr8 bacteriopha
38	89	5.3	595	16	Q9CFS9	Q9cfs9 lactococcus
39	89	5.3	672	5	Q95QP7	Q95qp7 caenorhabdi
40	89	5.3	1073	16	Q9A4M6	Q9a4m6 caulobacter
41	88.5	5.3	283	17	O26933	O26933 methanother
42	88.5	5.3	301	6	Q95MM0	Q95mm0 lepus capen
43	88.5	5.3	450	2	P74849	P74849 salmonella
44	88.5	5.3	466	16	Q97L19	Q97l19 clostridium
45	88.5	5.3	491	5	Q9XUF6	Q9xuf6 caenorhabdi

ALIGNMENTS

RESULT 1
Q9AQQ8
ID Q9AQQ8 PRELIMINARY; PRT; 320 AA.
AC Q9AQQ8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN-GLUTAMINASE PRECURSOR.
GN PRGA.
OS Chryseobacterium proteolyticum.
OC Bacteria; CFB group; Flavobacteria; Flavobacteriaceae;
OC Chryseobacterium.
OX NCBI_TaxID=118127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9670;
RX MEDLINE=21153247; PubMed=11231294;
RA Yamaguchi S., Jeenes D.J., Archer D.B.;
RT "Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme
RT that deamidates glutaminyl residues in proteins: purification,
RT characterization and gene cloning.";
RL Eur. J. Biochem. 268:1410-1421(2001).
DR EMBL; AB046594; BAB21508.1; -.
KW Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 136 320
SQ SEQUENCE 320 AA; 35044 MW; C67823D2BC131410 CRC64;

Query Match 100.0%; Score 1679; DB 2; Length 320;
Best Local Similarity 100.0%; Pred. No. 9.8e-125;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMI 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMI 60

Qy 61 KVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRY 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db     61 KVSFMLTAQFYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRY 120

Qy    121 FKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYAR 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 FKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYAR 180

Qy    181 AHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AHKMRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPS 240

Qy    241 LFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 LFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNSYLYDNNLINTNCVLTK 300

Qy    301 FSLLSGCSPSPAPDVSSCGF 320
          ||||||||||||||||||||
Db    301 FSLLSGCSPSPAPDVSSCGF 320


RESULT   2
P70004
ID   P70004      PRELIMINARY;      PRT;   342 AA.
AC   P70004;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ALPHA1-MICROGLOBULIN/BIKUNIN (AMBP) PRECURSOR.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Kawahara A., Hikosaka A., Sasado T., Hirota K.;
RT   "Thyroid hormone-dependent repression of alpha1-microglobulin/bikunin
RT   (AMBP) gene expression during amphibian metamorphosis.";
RL   Dev. Genes Evol. 206:355-362(1997).
DR   EMBL; D87752; BAA13453.1; -.
DR   HSSP; P02760; 1BIK.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   InterPro; IPR002345; Lipocalin.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF00014; Kunitz_BPTI; 2.
DR   Pfam; PF00061; lipocalin; 1.
DR   PRINTS; PR00759; BASICPTASE.
DR   PRINTS; PR00179; LIPOCALIN.
DR   SMART; SM00131; KU; 2.
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 2.
DR   PROSITE; PS00213; LIPOCALIN; 1.
KW   Serine protease inhibitor; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    342       POTENTIAL.
SQ   SEQUENCE   342 AA;  38558 MW;  3050508BFDC5C47B CRC64;


  Query Match             6.5%;  Score 108.5;  DB 13;  Length 342;
  Best Local Similarity   23.1%;  Pred. No. 1;
  Matches   49;  Conservative   30;  Mismatches   58;  Indels   75;  Gaps   12;

Qy     67 TAQFY----EIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDV---- 118
          | : |    :::||  :|     | | :  :|   | : ||: |       || ::
Db    145 TVKLYGRSPDLRPTLVDEFRQFALAQGIPEDS---IVMLPNNGE------CSPGEIEVRP 195

Qy    119 -RYFKTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCI--------- 168
           |  : :| :| :|   :           || |:: | :||  : || ||:
Db    196 RRTQRAVLPEEEEGSGME----------NSPFSKNKGESCRLAPASGPCLGNHNRYFYN 244

Qy    169 -------TFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYG-------NLKASTGTCCVA 214
                ||:|   ||            || :  ||: ::         |  : | |  |
Db    245 SSTMACETFQY--GGCLGN---------NNNFHSEKECLHDCRTEAACRLPITPGPCKTA 293

Qy    215 ---WSYHVA----ILVSYKNASG-----VTEK 234
              |:: |    :  ||    |      |||
Db    294 KTHWAFDAAQGKCVTFSYGGCQGNGNQFYTEK 325


RESULT   3
O02364
ID   O02364      PRELIMINARY;      PRT;   927 AA.
AC   O02364; P90955; O62351;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 102.3 KDA PROTEIN T01D3.6 IN CHROMOSOME V.
GN   T01D3.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A
RC   STRAIN=BRISTOL N2;
RA   Steward C.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS; ISOFORM T01D3.6A (SHOWN HERE) AND
CC       ISOFORM T01D3.6B; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
DR   EMBL; Z81527; CAB04280.1; -.
DR   EMBL; Z81110; CAB04280.1; JOINED.
DR   EMBL; Z81110; CAB03262.1; -.
DR   EMBL; Z81527; CAB03262.1; JOINED.
DR   EMBL; Z81527; CAB04279.1; -.
DR   EMBL; Z81110; CAB04279.1; JOINED.
DR   EMBL; Z81110; CAB03263.1; -.
DR   EMBL; Z81527; CAB03263.1; JOINED.
DR   HSSP; P02671; 1FZD.
DR   WormPep; T01D3.6A; CE12964.
DR   WormPep; T01D3.6B; CE18164.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR002919; TIL.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00147; fibrinogen_C; 1.
DR   Pfam; PF01826; TIL; 1.
DR   Pfam; PF00094; vwd; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 2.
DR   SMART; SM00186; FBG; 1.
DR   SMART; SM00216; VWD; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Alternative splicing; Calcium-binding; EGF-like domain; Glycoprotein;
KW   Hydroxylation; Hypothetical protein; Repeat.
FT   VARSPLIC    405    416       MISSING (IN ISOFORM T01D3.6B).
SQ   SEQUENCE   927 AA;  103413 MW;  71AEF0A61FC2B266 CRC64;


  Query Match             6.4%;  Score 108;  DB 5;  Length 927;
  Best Local Similarity   19.2%;  Pred. No. 3.8;
  Matches   85;  Conservative   47;  Mismatches  101;  Indels  210;  Gaps   24;

Qy     35 LSVNDSKLKDFGKTVPV----------GIDEENGMIKVSFMLTAQFYEIKPTKENEQYI 83
          |:  :  || :|    |           |:: |:   | : :     |:    :::::|:
Db     15 LAAAEDDLKCYGADCEVPHYMDDTEQCSGLECES---KETLLAMISHYQ---KEKHDEYL 68

Qy     84 GMLRQAVKNESPVHIFLKPNSNEIG----KVESASPE----------------DVR--- 119
            | : :::|| |  ||    ||:        : ||:                 | |
Db     69 APLTKPIRDESAVRNFL----NEVAVTGDLAATVSPQCSICSQPGLCNSGQCVPDARFPW 124

Qy    120 -YFKTILTKEVKGQTNKLASVIPDVATLNSLFNQI--KNQSCGTSTASSPCITFRYPVDG 176
```

```
           ||                 | || |:     |:|  |: ||| :
Db    125 QYF---------------YCVCPDYASGRFCQNEIKCKDNSCGKNA------------D 156

Qy    177 CYARAHKMRQI------LMNNGYDCEKQ-----------FVYGNLKAS-TGTCCVAWS- 216
          ||   |::  |         || ||: :             |  |:    |||   :|
Db    157 CYVANHQLNCICKPGYTARRNGRDCDMKVQQACMSGDPHYVTYDGLRFDYQGTCPYVFSQ 216

Qy    217 ------------------------YHVA----ILVSYKN------------ASGVTEKR 235
                                  ||::    : |   |              :||   :
Db    217 PCTTLPAPYLWYSVRAKNELPGKGYHISQVSEVEVDLHNLTIHVDGRSKTALVNGV---Q 273

Qy    236 IIDPSLFS-----------SGP-----------VTDTAWRNACV-----------NTSCG 262
          :: | |              ||            || | : : ||             | ||
Db    274 VLTPWYFPNKNTWTVRVRFSGSTFTIENDQGVVVTFTTYNSLCVQVPDIPEFNGATTLCG 333

Qy    263 SA------SVSSYANTAGNVY----YRSPSN-----------SYLYDNNLI--------- 292
           |       :    |  |:|      | | |           ::: |  ||
Db    334 LAGNIDGKKLDDVVNKNGSVLAIKSSRQPENNNHADFMKTEDTWITDKFLILRPGQENCI 393

Qy    293 -------NTNCVLTKFSLLSGCS 308
                |||||  |  ||   |:
Db    394 NGQTLDNNTNCVSTSISLAQSCA 416


RESULT   4
Q96SC3
ID   Q96SC3      PRELIMINARY;      PRT;  2673 AA.
AC   Q96SC3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FIBULIN-6 (FRAGMENT).
GN   FIBL-6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RA   Kostka G., Timpl R.;
RT   "Partial sequence of fibulin-6 with a c-terminal region related to
RT   domain II and III of the fibulin family.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ306906; CAC37630.1; -.
FT   NON_TER      1     1
SQ   SEQUENCE   2673 AA;  291017 MW;  BEAEC30B8340E272 CRC64;


  Query Match             6.1%;  Score 102;  DB 4;  Length 2673;
  Best Local Similarity   19.8%;  Pred. No. 43;
  Matches   66;  Conservative   58;  Mismatches  125;  Indels   84;  Gaps   16;

Qy     27 QEINGKEKLSVNDSKLKDFGK-----TVPVGIDEENGMIKV--------------SFML 66
          | : | | | :: ::::| |:     : | | |::  :::|              : :
Db    614 QTLGGGEVLRISTAQVEDTGRYTCLASSPAGDDDKEYLVRVHVPPNIAGTDEPRDITVLR 673

Qy     67 TAQF-YEIKPTKENEQYIGMLRQAVKNESPVHI-------FLKPNSNEIGKVESASPEDV 118
            |   | |         |  ||   : ::   :        :|: |: ::|       :
Db    674 NRQVTLECKSDAVPPPVITWLRNGERLQATPRVRILSGGRYLQINNADLG-------DTA 726

Qy    119 RYFKTILTKEVKGQTNK----LASVIPDV----ATLNSLFNQIKNQSCGTSTASSPCITF 170
           |  | :   : |:| :      :| |::     :|  | |:     |      :| ||:
Db    727 NY--TCVASNIAGKTTREFILTVNVPPNIKGGPQSLVILLNKSTVLECIAEGVPTPRITW 784

Qy    171 RYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGT-CCVAWSYHVAILVSYKNAS 229
          |   ||      :  | :: ||:      :        ||   |:|             ||:
Db    785 R--KDGAVLAGNHARYSILENGF----LHIQSAHVTDTGRYLCMA-----------TNAA 827

Qy    230 GVTEKRIID-----PSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNS 284
          | |::| ||      |   : ||   |   |      :| :  :              || :
Db    828 G-TDRRRIDLQVHVPPSIAPGPTNMTVIVNVQTTLACEATGIP------------KPSIN 874
```

```
Qy    285 YLYDNNLINTNCVLTKFSLLSGCS---PSPAPD 314
          :  : :|:| :      : |||  |    ||: |
Db    875 WRKNGHLLNVDQNQNSYRLLSSGSLVIISPSVD 907


RESULT   5
Q96RW7
ID   Q96RW7      PRELIMINARY;      PRT;  5636 AA.
AC   Q96RW7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HEMICENTIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Trent J.;
RT   "Human hemicentin gene.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF156100; AAK68690.1; -.
SQ   SEQUENCE   5636 AA;  613660 MW;  F000B319CED7B52C CRC64;


  Query Match             6.1%;  Score 102;  DB 4;  Length 5636;
  Best Local Similarity   19.8%;  Pred. No. 1.1e+02;
  Matches   66;  Conservative   58;  Mismatches  125;  Indels   84;  Gaps   16;

Qy     27 QEINGKEKLSVNDSKLKDFGK-----TVPVGIDEENGMIKV--------------SFML 66
          | : | | | :: ::::| |:     : | | |::  :::|              : :
Db   3577 QTLGGGEVLRISTAQVEDTGRYTCLASSPAGDDDKEYLVRVHVPPNIAGTDEPRDITVLR 3636

Qy     67 TAQF-YEIKPTKENEQYIGMLRQAVKNESPVHI-------FLKPNSNEIGKVESASPEDV 118
            |   | |         |  ||   : ::   :        :|: |: ::|       :
Db   3637 NRQVTLECKSDAVPPPVITWLRNGERLQATPRVRILSGGRYLQINNADLG-------DTA 3689

Qy    119 RYFKTILTKEVKGQTNK----LASVIPDV----ATLNSLFNQIKNQSCGTSTASSPCITF 170
           |  | :   : |:| :      :| |::     :|  | |:     |      :| ||:
Db   3690 NY--TCVASNIAGKTTREFILTVNVPPNIKGGPQSLVILLNKSTVLECIAEGVPTPRITW 3747

Qy    171 RYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGT-CCVAWSYHVAILVSYKNAS 229
          |   ||      :  | :: ||:      :        ||   |:|             ||:
Db   3748 R--KDGAVLAGNHARYSILENGF----LHIQSAHVTDTGRYLCMA-----------TNAA 3790

Qy    230 GVTEKRIID-----PSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPSNS 284
          | |::| ||      |   : ||   |   |      :| :  :              || :
Db   3791 G-TDRRRIDLQVHVPPSIAPGPTNMTVIVNVQTTLACEATGIP------------KPSIN 3837

Qy    285 YLYDNNLINTNCVLTKFSLLSGCS---PSPAPD 314
          :  : :|:| :      : |||  |    ||: |
Db   3838 WRKNGHLLNVDQNQNSYRLLSSGSLVIISPSVD 3870


RESULT   6
Q9HYT4
ID   Q9HYT4      PRELIMINARY;      PRT;   551 AA.
AC   Q9HYT4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HYPOTHETICAL PROTEIN PA3310.
GN   PA3310.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
```

```
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004753; AAG06698.1; -.
DR   InterPro; IPR003371; DUF146.
DR   Pfam; PF02418; DUF146; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   551 AA;  61421 MW;  3E3EAF5C308BE815 CRC64;


  Query Match             6.0%;  Score 101.5;  DB 16;  Length 551;
  Best Local Similarity   20.1%;  Pred. No. 6.5;
  Matches   64;  Conservative   40;  Mismatches   95;  Indels  119;  Gaps   15;

Qy      58 GMIKVSFMLTAQFYE--------------IKPTKENEQYIGMLRQAVKNESPV--HIFL 100
           | : :  ::   ||                | ||     ||  : :  |    :  |: :
Db     163 GCLLIVAVMVGSFYSTYAPIFREEDKLTHFINPT----NYIYAVGKYAKQRLVIKEHLVV 218

Qy     101 KPNSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNSL---------- 149
           :|    ||     :|  ::  |  |   | |:| :      |  :||
Db     219 EP----IGLDARQAPAALQRQKKSLLVFVVGETAR-----ADHFSLNGYARETNPELSKQ 269

Qy     150 ----FNQIKNQSCGTSTA-SSPCITFRYPVDGC---YARAHK----------MRQILMNN 191
               | |::  |||||| | ||:  :|| :      |: |:          :: : : |
Db     270 DIVNFTQVR--SCGTSTAVSVPCMFSQYPREDYSDKKAKTHEGLLDILQRAGVQVLWLEN 327

Qy     192 GYDCEKQFVYGNLKASTGTC--------------------CVAWSYHVAILVSYKNASG 230
             ||:            |||                   |:  |  | :    :  |
Db     328 NSDCK-----------GTCLRVPNRDIPKTQPSPFCDGKNCLDESLLVGL---QEYIDG 372

Qy     231 VTEKRII--------DPSLFSSGPVTDTAWRNACVNTSCGSASVSSYANTAGNVYYRSPS 282
           : :  ||         |  :   |     ::  |     || |     |
Db     373 LQDDAIIVLHSDGSHGPEYYERYPKEMERFQPVCRTNQLGSCSKEELVN---------- 421

Qy     283 NSYLYDNNLINTNCVLTK 300
             :||| :: |:  |||
Db     422 ---VYDNTILYTDHFLTK 436


RESULT   7
Q27167
ID   Q27167      PRELIMINARY;      PRT;  2395 AA.
AC   Q27167;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   51B TYPE SURFACE PROTEIN.
OS   Paramecium tetraurelia.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC   Paramecium.
OX   NCBI_TaxID=5888;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOCK 51;
RX   MEDLINE=95098630; PubMed=7800503;
RA   Scott J., Leeck C., Forney J.;
RT   "Analysis of the micronuclear B type surface protein gene in
RT   Paramecium tetraurelia.";
RL   Nucleic Acids Res. 22:5079-5084(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOCK 51;
RA   Scott J.M., Leeck C.L., Forney J.D.;
RL   Genetics 133:189-198(1993).
DR   EMBL; U07603; AAA81947.1; -.
DR   EMBL; L04795; AAA16710.1; -.
DR   InterPro; IPR002895; Paramecium_SA.
DR   InterPro; IPR003659; PSI.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF01508; Paramecium_SA; 28.
DR   SMART; SM00423; PSI; 3.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
SQ   SEQUENCE   2395 AA;  246430 MW;  8632E022649B2D61 CRC64;


  Query Match             6.0%;  Score 101;  DB 5;  Length 2395;
  Best Local Similarity   22.3%;  Pred. No. 45;
  Matches   62;  Conservative   22;  Mismatches   94;  Indels  100;  Gaps   12;

Qy     122 KTILTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVD------ 175
           ||         || |        |:||      :::|  | | : |       |
Db     464 KTCANAPSTNTTNDLCVAFLSSCTVNSTNAGCVDKTCENSLAQTIC-------DKDLNNK 516

Qy     176 ------GCYARAHKMRQILMNNGYDCEKQFVYGNLKASTGTCCV-------AWSYHVAIL 222
                 || |   :       ||:   | |   ::||| ||       | :   |
Db     517 ACIWKGKCYKRECVLASSTTATHADCQTYDV-GCTLSNTGTGCVPLPLKCEAITIEAACN 575

Qy     223 VSYKNASGV----------------------TEKRIIDPSLFSSG------------- 245
           :  :  |||                      |     |   : ||
Db     576 IRLQVTSGVKSYPLCGWNGSSCIDKACSTAPKTTATTSDCGTYKSGCVANNPVNGSIQGC 635

Qy     246 ---PVTDTA-----------------WR---NACVNTSCGSASVSSYANTAGNVYYRSP 281
               | |  |                 |    :|||  || :|||::   | | :   |
Db     636 QDLPTTCAARKSTENCEITRTGFPTCLWNSATSACVEKSCSTASVTT---TTGFLTVFSN 692

Qy     282 SNSYLYDNNLINTNCVLTKFSLLSGCSPSPAPDVSSCG 319
           :|   |   | |: |:    :   || | |    ||||
Db     693 TNCLAY---LSNSACIAN--NTADGCIPKP----SSCG 721


RESULT   8
O40366
ID   O40366      PRELIMINARY;      PRT;  848 AA.
AC   O40366;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRMNR;
RX   MEDLINE=98216834; PubMed=9557756;
RA   Blanchard A., Ferris S., Chamaret S., Guetard D., Montagnier L.;
RT   "Molecular evidence for nosocomial transmission of human
RT   immunodeficiency virus from a surgeon to one of his patients.";
RL   J. Virol. 72:4537-4540(1998).
DR   EMBL; U85919; AAC59357.1; -.
DR   InterPro; IPR000328; Env_GP41.
DR   InterPro; IPR000777; GP120.
DR   Pfam; PF00516; GP120; 1.
DR   Pfam; PF00517; GP41; 1.
KW   AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT   NON_TER       1      1
SQ   SEQUENCE   848 AA;  96100 MW;  4C224022C395F34B CRC64;


  Query Match             6.0%;  Score 100;  DB 15;  Length 848;
  Best Local Similarity   21.2%;  Pred. No. 15;
  Matches   60;  Conservative   39;  Mismatches   94;  Indels   90;  Gaps   13;

Qy     101 KPNSNEIGKVESASPEDVRYFKT----ILTKEVKGQTNKLASVI--PDVATL-------- 146
           | |||     ||   || | |     |     :| |:: :| |: |:    ||  |
Db     133 KSNSN---TTESPPIEDPRREMTNCSFNMTTELRDKTKKVYSLFYKLDVVQLDEENSNSS 189

Qy     147 NSLFNQIKNQSCGTSTASSPC--ITFR-YPVDGCYARAHKM------------------ 184
           |  ::  :  :| ||| :  |  ::|   |:  |     : :
Db     190 NGNYSYYRLVNCNTSTITQACPKVSFEPIPIHYCAPAGYAILKCKDKNFNGKGPCSNVST 249
```

```
Qy    185 ------------RQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVT 232
                      |:|:|     ||  :    |: : |  :    |  : ::    :  |
Db    250 VQCTHGIKPVVSTQLLLNGSLAEEKIVIRSKNKSDSTTSIIVQLVH-PVKINCTRPNNNT 308

Qy    233 EKR-------------IIDPSLFSSGPVTDTAWRNACVN-----------------TSC 261
           |               ||      :|  |  | |||  :|                  :|
Db    309 RKSVRIGPGQAFYTRDIIGDIRQASCTVNKTEWRNTLLNVSEQLKEYFNKNKATFNRSSG 368

Qy    262 GSASVSSYA-NTAGNVYYRSPS----NSYLYDNNLINTNCVLT 299
          |   ::::: |  |  :| : |    |: |:| |  |||  :|
Db    369 GDLEITTHSFNCRGEFFYCNTSGLFNNTLLHDTN--NTNGNIT 409


RESULT   9
O94317
ID   O94317      PRELIMINARY;      PRT;   534 AA.
AC   O94317;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 53.4 KDA SERINE-RICH PROTEIN C215.13 IN CHROMOSOME II
DE   PRECURSOR.
GN   SPBC215.13.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- SIMILARITY: TO YEAST STA1.
DR   EMBL; AL033534; CAA22127.1; -.
KW   Hypothetical protein; Glycoprotein; Signal.
FT   SIGNAL        1     14       POTENTIAL.
FT   CHAIN        15    534       HYPOTHETICAL SERINE-RICH PROTEIN C215.13.
FT   CARBOHYD     31     31       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    426    426       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   534 AA;  53361 MW;  D62DC15C80A698E4 CRC64;


  Query Match             5.9%;  Score 99.5;  DB 3;  Length 534;
  Best Local Similarity   22.7%;  Pred. No. 9;
  Matches   55;  Conservative   39;  Mismatches  123;  Indels   25;  Gaps    6;

Qy     85 MLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVA 144
          :: |   | |   | |        : |    :::||:      |  |        |  :
Db     24 LVHQIYGNSSFTKISLNQLEGRDSQEELQRRQEIRYYGR--AAETGGTPTYYGYATPTSS 81

Qy    145 TLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARAHKMRQILMNNGYDCEKQFVYGNL 204
          :  |:|::    |  |:: |||  ::  |            |:  :  :       |
Db     82 SEPSIFSESATPS-ETNSYSSPVSSYSDPATS---------QLPSSTSFFSPTSSEYTPS 131

Qy    205 KASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVNTSCGSA 264
             : :     |   ||| |    |   |   |   || || |:| : : :   :||
Db    132 STESSSLLDPSSVSSAILPS----STSVEVSISSSSLSSSDPLTSSTFSSLSSSTSSSQP 187

Qy    265 SVSS-YANTAGNVYYRSPSNSYLYDNNLI--------NTNCVLTKFSLLSGCSPSPAPDV 315
          ||||  ::|  :    | |:||| ::::        :::  ||  || :   || :
Db    188 SVSSTSSSTFSSAAPTSTSSSYLSSSSVVSSSSSPSSSSSSTLTSSSLSTSSIPSTSSSS 247

Qy    316 SS 317
          ||
Db    248 SS 249


RESULT  10
Q9S746
ID   Q9S746      PRELIMINARY;      PRT;   594 AA.
AC   Q9S746;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ACE (AT1G72970/F3N23_17).
GN   ACE OR F3N23.17.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Araki T., Nakatani-Goto M.;
RT   "ADHESION OF CALYX EDGES (ACE), cDNA.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA   Altafi H., Araujo R., Huizar L., Rowley D., Chen S., Harman P.,
RA   Hicks R., Huerta M., Mason S., Siepel J., Zimmerman M., Buehler E.,
RA   Dunn P., Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J.,
RA   Liu S., Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA   Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA   Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA   Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA   Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA   Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA   Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB027458; BAA77837.1; -.
DR   EMBL; AC008017; AAD55644.1; -.
DR   EMBL; AY054193; AAL06854.1; -.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR000172; GMC_oxred.
DR   Pfam; PF00732; GMC_oxred; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00624; GMC_OXRED_2; UNKNOWN_1.
SQ   SEQUENCE   594 AA;  65344 MW;  D6A8682FBBF73881 CRC64;


  Query Match             5.9%;  Score 99;  DB 10;  Length 594;
  Best Local Similarity   21.0%;  Pred. No. 11;
  Matches   49;  Conservative   28;  Mismatches   74;  Indels   82;  Gaps   11;

Qy     81 QYIGMLRQAVK-------NESP----VHIFLKPNSNEIGKVESASPEDVRYFKTILTKEV 129
          | :|: :  |          :||      |  :  | ||:             | ||  |:
Db    353 QTVGITKMGVYVEASTGFGQSPESIHTHYGIMSNKNEL------------FSTIPAKQR 399

Qy    130 KGQTNKLASVIPDVATLNSLFN----------------QIKNQSCGTSTASSPCITFRY 172
          : :  : | :   :   |:  ||                : |    |:   :| :|| |
Db    400 RPEATQ-AYITRNKYQLHEAFNGSFILEKLAYPISRGHLSLVN----TNVDDNPSVTFNY 454

Qy    173 ---PVD--GCYARAHKMRQILMNNGY----DCEKQFVYGNLKAST--------------- 208
             |||   |       : ::: :| :     |:|| |:  |  |
Db    455 FKHPVDLQRCVEAIRLVSKVVTSNRFLNYTQCDKQNVHKMLSLSVKANINLRPKQLNDTK 514

Qy    209 -------GTCCVAWSYHVAILVS-----YKNASGVTEKRIIDPSLFSSGPVTD 249
                  |    | ||   ||       :   ||   |:|| | |   | |:
Db    515 SMAQFCKDTVVTIWHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFDESPGTN 567


RESULT  11
Q9KB36
ID   Q9KB36      PRELIMINARY;      PRT;   419 AA.
```

```
AC   Q9KB36;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BH2099 PROTEIN.
GN   BH2099.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
DR   EMBL; AP001514; BAB05818.1; -.
KW   Complete proteome.
SQ   SEQUENCE   419 AA;  46289 MW;  50B84D6C23EAE2C4 CRC64;
```

Query Match 5.9%; Score 98.5; DB 16; Length 419;
Best Local Similarity 23.0%; Pred. No. 8;
Matches 65; Conservative 38; Mismatches 99; Indels 81; Gaps 12;

```
Qy      12 VTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFMLTAQFY 71
Db      30 VIVLNFAFAGLFGGDQELNLDLTLAV------------VNEDDETGTI-------AQLQ 69

Qy      72 E--IKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTI----- 124
Db      70 EKLIKEASFKETEVARLVEQVGDIQPIPLLF----------EYLSSEDLQDWLTVYELTE 119

Qy     125 ---LTKEVKGQTNKLASVIPDVATLNSLFNQIKNQSCGTSTASSPCITFRYPVDGCYARA 181
Db     120 AEAVAKVKEGAADGIL-VIPDGFTVNSLYAAL----VGEPTASS----LIYKVEMETSNS 170

Qy     182 HKMRQIL------MNNGYDCEKQFVYGNLKASTG------------------TCCVAWS 216
Db     171 ATLYQIIDGFLDHMNNHFALQQMGEVQEANAPEGGLEKIGVGEAFTLTQYFTISMSALFA 230

Qy     217 YHVAILVSYKNASGVTEK---RII----DPSLFSSGPVTDTAW 252
Db     231 LFLAATVATKTGEEIRQKVFHRILLTDRKPILFLVGKIVSTAF 273
```

```
RESULT  12
O26459
ID   O26459      PRELIMINARY;      PRT;   534 AA.
AC   O26459;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CONSERVED PROTEIN.
GN   MTH359.
OS   Methanothermobacter thermautotrophicus.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000821; AAB84865.1; -.
DR   InterPro; IPR002931; Transglut_core.
DR   Pfam; PF01841; Transglut_core; 1.
DR   SMART; SM00460; TGc; 1.
KW   Complete proteome.
SQ   SEQUENCE   534 AA;  57628 MW;  588412BD8D342E26 CRC64;
```

Query Match 5.8%; Score 97; DB 17; Length 534;
Best Local Similarity 20.9%; Pred. No. 14;
Matches 57; Conservative 34; Mismatches 108; Indels 74; Gaps 9;

```
Qy      44 DFGKTVPVGI----------DEENGMIKVSFMLTAQ----FYEIKPTKENEQYIGMLRQ 88
Db     277 NLGKNTPVAPVTAGEAANPAGDARGTLTRLEYLKVASGVQGFMEKNHRAPN--YAGSSKG 334

Qy      89 AVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTILTKEVKGQTNKLASVIPDVATLNS 148
Db     335 RISYESLVY--------SISRVLSFYSSNKR------------------LPNTVTVTR 366

Qy     149 LFNQIKNQS-----CGTSTASSPCITFRYPVD----------------GCYARAHK---- 183
Db     367 LASSLKNRPENDPYRGESTARYLASSASCPVDSPDIRSLASEITRGLTSTFSRAEAVFGW 426

Qy     184 MRQILMNNGYDCEKQFVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFS 243
Db     427 VRDNINYSFYYNTKYGAVGTLKNRTGNCV----DHTHLLVALARAAGIPARYVHGTCNFT 482

Qy     244 SGPVTDTAWRNACVNTSCGSASVSSYANTAGNV 276
Db     483 SGNVYGHVWAQLLVGDTWYAADATSSRNSLGVV 515
```

```
RESULT  13
O72736
ID   O72736      PRELIMINARY;      PRT;   564 AA.
AC   O72736;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   A54R PROTEIN.
GN   A54R.
OS   Cowpox virus (CPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10243;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RX   MEDLINE=98229462; PubMed=9568042;
RA   Shchelkunov S.N., Safronov P.F., Totmenin A.V., Petrov N.A.,
RA   Ryazankina O.I., Gutorov V.V., Kotwal G.J.;
RT   "Species-specific differences in genome organization of cowpox,
RT   smallpox, and vaccinia viruses.";
RL   Virology 243:432-460(1998).
DR   EMBL; Y15035; CAA75274.1; -.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR001798; Kelch.
DR   Pfam; PF00651; BTB; 1.
DR   Pfam; PF01344; Kelch; 6.
DR   PRINTS; PR00501; KELCHREPEAT.
DR   SMART; SM00225; BTB; 1.
DR   PROSITE; PS50097; BTB; 1.
SQ   SEQUENCE   564 AA;  64825 MW;  EF268497F8DC6AF4 CRC64;
```

Query Match 5.8%; Score 97; DB 12; Length 564;
Best Local Similarity 21.8%; Pred. No. 15;
Matches 72; Conservative 52; Mismatches 108; Indels 98; Gaps 18;

```
Qy      11 FVTVLTFNSCADSNGNQEIN--GKEKLSVNDSKLKDFGKTVPVGIDEENGMIKVSFML-T 67
           : ::|  |:  |||   |:|    :  ||||  | |:   :|: :  :|    | ::| |
Db      45 YFSILFSNNFIDSN-EYEVNLSHLDYQSVND--LIDYIYGIPLSLTNDN----VKYILST 97

Qy      68 AQFYEI-KPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPEDVRYFKTIL- 125
           | | :|     | |:|| :    :|    :|:    :|:  |:|||      | |||
Db      98 ADFLQIGSAITECEKYI-LKNLCSRNCIDFYIYADKYNNK--KIESAS------FNTILR 148

Qy     126 --TKEVKGQTNKLASVIPDVATLNSLFNQIKN----------------QSCGTS------ 161
             : :  :  |  :    :  |:     |||                 |||
Db     149 NILRLINDENFKYLTEESMIKILSDDMLNIKNEDFAPLILIKWLESTQQSCTVELLRCLR 208

Qy     162 -TASSP-----------------CITFRYPV---DGCYARAHKMR--QILMNNGYDCEKQ 198
            :  ||                 ||||   :   |  : | | :    | ::| :|
Db     209 ISLLSPQVIKSLYSHQLVSSIYECITFLNNIAFLDESFPRYHSIELISIGISNSHDKISI 268

Qy     199 FVYGNLKASTGTCCVAWSYHVAILVSYKNASGVTEKRIIDPSLFSSGPVTDTAWRNACVN 258
           | | | :|     :  |  :  |:           ::|  ::  |    : :|:: |
Db     269 NCY-NHKKNTWEMISSRRYRCSFAVA----------VLDNIIYMMGGYDQSPYRSSKV- 315

Qy     259 TSCGSASVSSYANTAGNVYYRSPSNSYLYD 288
                       : | : :||::||
Db     316 -----------------IAYNTCTNSWIYD 328


RESULT  14
Q99RZ3
ID   Q99RZ3      PRELIMINARY;      PRT;   260 AA.
AC   Q99RZ3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MODA PROTEIN.
GN   MODA OR SA2074.
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Staphylococcus.
OX   NCBI_TaxID=158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
RA   Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
RA   Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA   Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA   Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA   Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003136; BAB43371.1; -.
DR   HSSP; P37329; 1WOD.
KW   Complete proteome.
SQ   SEQUENCE   260 AA;  29117 MW;  7A5D4A01A4482C4D CRC64;


  Query Match             5.7%;  Score 96.5;  DB 16;  Length 260;
  Best Local Similarity   22.5%;  Pred. No. 6.3;
  Matches   42;  Conservative   39;  Mismatches   59;  Indels   47;  Gaps    9;

Qy       1 MKNLFLSMMAFVTVLTFNSCADSNGNQEINGKEKLSVNDSKLKDF--GKTVPVGIDEENG 58
           ||     :||   ||    |::|| |           |:|| |   ||   : :
Db       3 MKRFIAIVMALFLVLA--GCSNSNDN----------NESKKDDADNGKKQEIQVAAAAS 49

Qy      59 MIKVSFMLTAQF---YEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIG--KVESA 113
           :  |:  | ::|   ::    | |    | ||: ::: :|| :|:  |: ::   | ::
Db      50 LTDVTKKLASEFKKEHKNADIKFNYGGSGALRKQIESGAPVDVFMSANTKDVDALKDKNK 109

Qy     114 SPEDVRYFKTIL----------TKEVKGQTNKLASVIPDVATL----------NSL 149
           : :  :| |  |            |::|   :|||  : :| |:           |:|
Db     110 AHDTYKYAKNSLVLIGDKDSNYTSVKDLK-DNDKLA--LGEVKTVPAGKYAKQYLDNNNL 166

Qy     150 FNQIKNQ 156
           | :::::
Db     167 FKEVESK 173


RESULT  15
Q9U459
ID   Q9U459      PRELIMINARY;      PRT;  5458 AA.
AC   Q9U459; Q9GPG4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ERYTHROCYTE MEMBRANE-ASSOCIATED GIANT PROTEIN ANTIGEN 332.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCC1/HN;
RA   Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
RT   "Molecular cloning and structure analysis of the Plasmodium falciparum
RT   erythrocyte membrane-associated giant protein Ag332 (Pf332) gene.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202180; AAF15293.3; -.
DR   InterPro; IPR001313; PUM.
SQ   SEQUENCE   5458 AA;  615269 MW;  4F6A37D92BD80172 CRC64;


  Query Match             5.7%;  Score 96.5;  DB 5;  Length 5458;
  Best Local Similarity   22.6%;  Pred. No. 2.9e+02;
  Matches   36;  Conservative   32;  Mismatches   64;  Indels   27;  Gaps    5;

Qy      21 ADSNGNQEINGKEKLSVNDSKLKDFGKTVPVGIDEE--NGMIKVSFMLTAQFYEI--KPT 76
           |: :  :|| | |: : |:| :|       | :|:   |  |:     : : | :  |
Db    4654 AEVSVTEEIPGVEETTNNESHVKGENVVNEVVVDDNSVNDEIQFDDDSSIEIYTVDSKDV 4713

Qy      77 KENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVES-----------ASPEDVR----- 119
              |:     |: |:::  :||  | | | : |:|:             | |::
Db    4714 FHKEENYDSFREEVRSDENIHISRKKNENFVKKIENEKSSVGHVPTEYTSKENIAEEVPS 4773

Qy     120 --YFKTILT----KEVKGQTNKLASVIPDVATLNSLFNQ 152
             ||  :|    |||| : | :  |  :| :  ::  :
Db    4774 HIMFKENVTEEVPKEVKYEENTVEEVFEEVTSKENIIEE 4812


Search completed: June 27, 2002, 21:58:29
Job time: 662 sec
```